Investigating the Population Structure of Queensland Invasive *Streptococcus pneumoniae* Isolates in Children: Using a Modified Multi-Locus Variable number of tandem repeat Analysis and a Novel MinimumSNPs Capsular Typing Method

By

Rachael E. Rayner

B. App. Sc. (major Forensic Science) – with Distinction

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Institute of Health and Biomedical Innovation (IHBI), School of Biomedical Sciences

Faculty of Health

Queensland University of Technology

August 2015

Principal supervisor: Assoc. Prof. Flavia Huygens (QUT)

Associate supervisors: Prof. Louise Hafner (QUT)

Mr. John Savill (QHFSS)

KEYWORDS

Streptococcus pneumoniae

Invasive Pneumococcal Disease (IPD)

Epidemiology

Pneumococcal population

Multi-Locus Sequence Typing (MLST)

Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA)

Capsule typing

Genotyping

Minimum SNPs

Real-time PCR

ABSTRACT

Streptococcus pneumoniae is a Gram-positive bacterium that is associated with a range of diseases including pneumonia, otitis media and invasive pneumococcal disease (IPD) including meningitis and septicaemia. It is estimated that approximately 800 000 children under five die each year, worldwide, from pneumococcal diseases (O'Brien *et al.,* 2009). As well as this, *S. pneumoniae* causes high morbidity (estimated 14.5 million episodes of serious pneumococcal disease in 2000) and high economic burden globally, especially in underdeveloped countries (O'Brien *et al.,* 2009). As a result, a number of strategies to combat pneumococcal diseases have been implemented, including antibiotics and vaccines. Unfortunately, despite initial decreases in pneumococcal diseases, this bacterium has managed to overcome the effects of drugs and vaccines by developing resistance or evasive mechanisms, and hence pneumococcal diseases continue to persist throughout the world.

After the introduction of the first childhood vaccine, called Prevenar® or 7-valent pneumococcal conjugate vaccine (7vPCV; Wyeth Pharmaceuticals) in Australia during 2005, IPD was reported to decline as observed worldwide (Roche *et al.,* 2008). Unfortunately, due to a phenomenon called serotype replacement, incidences of IPD began to increase. Therefore an upgraded vaccine, Prevenar13® (13vPCV; Pfizer) was licensed in Australia in July 2011 (Selvey, 2011). Since then, there has been no published study on the genetic population of *S. pneumoniae* in Queensland. Surveying *S. pneumoniae* is important to detect changes in the population, particularly after the introduction of the vaccine.

Therefore, a number of genotyping methods have been used worldwide to study the pneumococcal population. Multi-Locus Sequence Typing (MLST) is considered the 'gold standard' genotyping method for *S. pneumoniae*. However, due to the fact that MLST is expensive and better suited for evolutionary studies rather than epidemiology studies, another genotyping method has been developed, called Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA). This new method has been shown to be cheaper, faster and highly discriminatory compared to MLST. However, current MLVA methods also pose limitations, which include failure to obtain complete genetic profiles for all pneumococcal strains, and the fact that there is no universally accepted method that has been tested worldwide. This study firstly aims to investigate the current MLVA methods and propose modifications to the protocol for improved genetic typing of *S. pneumoniae*. Comparisons of our modified MLVA4 method has been made against

MLST and two other published MLVA methods (Elberse *et al.,* 2011a; Van Cuyck *et al.,* 2012).

Characterisation of the Queensland pneumococcal population was performed as our second aim. A time frame from 2007 to 2012 was used, and *S. pneumoniae* isolates, taken from children under the age of 15 years, were genotyped using two MLVA methods (including the modified MLVA4 method) and MLST. The major findings, indicated that there were a number of clonal complexes (CC) of pneumococcus causing invasive disease in children, with the dominant sequence types (ST) and MLVA types (MT) including *S. pneumoniae* serotype 1 (ST306; MT36), 3 (ST1230; MT8), 7F (ST191; MT27) and 19A (ST411; MT59 and ST63; MT60). Following on from this, our third aim was to determine whether any changes of non-13vPCV serotypes could be detected. At this point there is no observable increase in non-13vPCV serotypes causing IPD, however these serotypes still cause approximately 33% of IPD in Queensland children.

Upon examination of the pneumococcal population in Queensland, it was observed that several pneumococcal strains appeared to have switched their protective capsule, the outer-most layer covering the cell wall. This polysaccharide capsule is thought to play a role in pneumococcal virulence, and there are at least 98 capsule types, (also called serotypes) known worldwide (Bentley *et al.,* 2006; Bratcher *et al.,* 2010; Calix & Nahm, 2010; Calix *et al.,* 2012; Ko *et al.,* 2013; Oliver *et al.,* 2013; Park *et al.,* 2007; Park *et al.,* 2015). The phenomenon of capsule switching is thought to occur when the bacteria take up naked DNA from the external environment through a process called horizontal transformation. This new DNA is incorporated into the genome, and a capsule switch occurs when capsular genes are taken up. The result is a transformed polysaccharide capsule which could enable the bacteria to evade the vaccine and potentially lead to vaccine escape strains.

Therefore, despite the widely used 'gold standard' capsular typing method, the serotyping/Quellung reaction, there are also a number of limitations. As a result, we proposed to develop a novel capsule typing method based on the use of a bioinformatics program called Minimum SNPs and molecular PCR methods. A number of molecular PCR methods for capsule typing have been published however they are limited to the most common capsule types associated with pneumococcal disease. We wanted to expand on these methods and develop a technique which could enable the ability to type any of the 98 known pneumococcal capsules. As well as this, due to potential capsule switching detected previously, we can combine this capsule typing method with the previous MLVA4 method to determine capsule switching events.

The major outcomes of this project include the development of a modified MLVA4 method and the development of a novel capsule typing method, which have both been applied to Queensland invasive *S. pneumoniae* isolates. A database of our MLVA4 profiles has been made available, and isolates with a MLST type has also been submitted to the open-access international MLST database (http://pubmlst.org/spneumoniae/). MLVA4 is highly discriminatory, and the application of MLVA to other bacterial and non-bacterial organisms is becoming more popular, particularly because of the cheaper costs and speed of the method. This has been particularly important for epidemiology studies of outbreaks and bioterrorism events.

Furthermore, the novel capsule typing method based on the Minimum SNPs bioinformatics program can characterise the pneumococcal capsule types in Queensland, although further optimisation of the method is required. The combination of the novel capsule typing with MLVA4 allows the identification of capsule switching events, of which a switch from serotype 19A to 15C and switch from serotype 1 to 4 has been observed in the Queensland pneumococcal population. Minimum SNPs bioinformatics program could also be applied for developing molecular capsule typing methods of other microorganisms, particularly since a number of Gram-negative, Gram-positive and even certain fungus carry a polysaccharide capsule.

Future directions for research include the increased collaboration to investigate or establish a universal MLVA genotyping method and database for genotyping *S. pneumoniae*. Continued monitoring of the pneumococcal population in Queensland is advised, particularly for at least five years after the introduction of the most recent vaccine in 2011 to gain insight into the effects of the vaccine and whether serotype replacement will occur. Future investigation into the novel capsule typing method can be performed, including the ability to multiplex the PCR reactions so that the protocol is faster and potentially cheaper, the addition of other capsule genes that identify capsulated or non-capsulated pneumococci, the ability to capsule type directly from clinical samples (e.g. blood, sputum, etc.) without the need to culture, and the verification of whether capsule switching can be confidently determined using this simple PCR method or whether whole genome sequence or capsule sequencing is required. The significance of this PhD thesis provides further insight into *S. pneumoniae* epidemiology studies and the interpretation of *S. pneumoniae* population structures using genotyping methods, and the information may impact future vaccine production and health preventative measures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

7vPCV	7-valent Pneumococcal Conjugate Vaccine
10vPCV	10-valent Pneumococcal Conjugate Vaccine
13vPCV	13-valent Pneumococcal Conjugate Vaccine
15vPCV	15-valent Pneumococcal Conjugate Vaccine
23PPV	23 Pneumococcal Polysaccharide Vaccine
AW	Adjusted Wallace coefficient
BHI	Brain Heart Infusion
CC	Clonal Complex
CI	Confidence Interval
CPS	Capsule Polysaccharide
CSF	Cerebrospinal fluid
CSP	Competent Stimulating Peptide
DI	Discriminatory Index (Simpson's Index of Discrimination)
DLV	Double Locus Variant
DNA	Deoxyribonucleic acid
dsDNA	double-stranded DNA
ESI-MS	Electrospray Ionisation – Mass Spectrometry
HBA	Horse Blood Agar
IPD	Invasive Pneumococcal Disease
IS	Insertion sequence
MDR	Multi-Drug Resistant
MLBT	Multi-Locus *boxB* Typing
MLEE	Multi-Locus Enzyme Electrophoresis
MLST	Multi-Locus Sequence Typing
MLVA	Multi-Locus Variable number tandem repeats Analysis
MM	Master Mix
MT	MLVA Type
MS	Mass Spectrometry
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NNDSS	National Notifiable Disease Surveillance System
PCR	Polymerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis

PMEN	Pneumococcal Molecular Epidemiology Network
PSGS	Pneumococcal Serotyping and Genotyping
QHFSS	Queensland Health Forensic and Scientific Services
QUT	Queensland University of Technology
rtPCR	real-time PCR
SLV	Single Locus Variant
SNP	Single Nucleotide Polymorphism
ssDNA	single-stranded DNA
ST	Sequence Type
TE	Tris Ethylenediaminetetraacetic-acid
TLV	Triple Locus Variant
URT	Upper Respiratory Tract
VNTR	Variable Number of Tandem Repeats

LIST OF DEFINITIONS

BOX loci – a type of VNTR loci, and is specifically identified as a 45-base pair tandem repeat located between *boxA* and *boxC* loci in *Streptococcus pneumoniae*

Capsule cassette – the collection of genes that code for the polysaccharide capsule in *S. pneumoniae*

Capsule switching – the process when pneumococci exchange capsule genes through horizontal transformation so that the expressed capsule (serotype) changes to another

Clonal Complex (CC) – a group of isolates that are genetically related either by single or double locus variants in the genetic profile

Competence - the stage during a microorganism's life where transformation occurs, indicating that it has the ability to take up DNA and transform

Cross protection – (also referred to as cross-immunity) occurs when the antibodies in the pneumococcal vaccine bind to more than one specific capsule type, providing protection against both capsule types

Emergent serotype - a non-vaccine targeted serotype that becomes a carrier in >5% of the population after the introduction of a vaccine, unless already exceeding this before introduction of a vaccine

Epidemiology – the study of patterns, effects and causes of disease and health conditions in a defined population

Fratricide - the killing of neighbouring pneumococcal bacteria by another pneumococcal bacterium

Genotype – the combination of alleles situated within the bacterial genome

Herd immunity – a form of immunity that occurs when a large population (or herd) is vaccinated to the point that it provides a measure of protection for individuals who have not developed immunity

Horizontal transformation – the transfer of genes from one bacterium to another via genetic manipulation of the cells so that the foreign genetic material can be expressed in the new cell

Invasive pneumococcal disease (IPD) – the isolation of *S. pneumoniae* from normally sterile body specimens such as blood, cerebral spinal fluid, pleural fluid and tissues

Minimum SNPs – a computer software program that traditional identifies SNPs within a genome, but has been utilised to identify highly variable genes within a cassette

Mother template – a manually engineered string of pneumococcal capsule genes that could be present in any pneumococcal serotype; used for creating 'pseudoDNA' sequences for the MinimumSNPs program

Multiplex – a chemical reaction designed to target more than one DNA region

Opsonophagocytosis - a mechanism of the host to clear out streptococcal infection via phagocytosis enhanced by an opsonin molecule

Pneumococcal conjugate vaccine (PCV) - a vaccine that conjugates a polysaccharide antigen to an immunogenic carrier protein so that it can elicit an immune memory response to protect against *S. pneumoniae*

PseudoDNA sequence – a mock DNA sequence that only contains A and T (A=absent and T=present) that is based on the pneumococcal capsule cassette and is used by the Minimum SNPs program to identify highly variable "SNPs" (each SNP position codes for a capsule gene)

Serogroup – designates a group of bacteria sharing a common antigen

Serotype – a distinct characterisation of microorganisms based on the specific antigens on the cell surface as determined by serological testing (e.g. Quellung reaction). This provides further differentiation within a serogroup.

Serotype replacement - the increase in non-vaccine targeted serotypes after a noticeable decrease in vaccine-targeted serotypes in a given population

Serotyping – a method to detect pneumococcal serotypes by using antibodies to bind and react to the pneumococcal polysaccharide capsule, causing it to become opaque and enlarged when visualised under a microscope

Singleplex – a chemical reaction that only targets one DNA region

Syntenic – when two or more genomic regions (such as the pneumococcal capsule cassette) derive from a common ancestral genomic region

Vaccine escape - to describe a pneumococcal strain that has increased in numbers in a given population after switching its capsule from a vaccine-targeted serotype to a non-vaccine-targeted serotype

VNTR loci – a sequence in a genome where a short nucleotide sequence is organized in a tandem repeat

LIST OF PUBLICATIONS

Journal Papers

Rayner RE., Savill J., Hafner LM., and Huygens F. (2015). Genotyping *Streptococcus pneumoniae*. Future Microbiology, 10(4): 653-664.

Rayner RE., Savill J., Hafner LM., and Huygens F. (2015). Modified MLVA for genotyping Queensland invasive *Streptococcus pneumoniae*. PLoS One, 10(4): e0121870. doi:10.1371/journal.pone.0121870.

Rayner RE., Savill J., Hafner LM., Huygens F. The changing population structure of invasive *S. pneumoniae* in Queensland children. Draft (to be submitted to PLoS One).

Conferences and Presentations

Rayner RE., Savill J., Hafner LM., Huygens F. A novel capsular typing method for *Streptococcus pneumoniae* using Minimum SNPs.

- Microbiology in Maleny – Queensland section of Australian Society of Microbiology (ASM), Annual scientific meeting, November 29, 2014, Maleny, Australia.
- IHBI Inspires Postgraduate Student Conference, Annual Scientific meeting, November 20 - 21, 2014, Gold Coast, Australia.
- XIX Lancefield International Symposium on Streptococci and Streptococcal Diseases (LISSSD), Scientific Meeting, November 9 -12, 2014, Buenos Aires, Argentina.

--Australian Society of Microbiology, Annual Scientific meeting, July 12 – 15, 2015, Canberra, Australia.

Rayner RE., Savill J., Hafner LM., Huygens F. The changing population structure of invasive pneumococci in Queensland children.

- Microbiology at QUT and Beyond, Scientific meeting. October 2014, Brisbane, Australia.

- Australian Society for Microbiology (ASM), Annual Scientific Meeting. July 2013, Adelaide, Australia.

Rayner RE., Savill J., Hafner LM., Huygens F. A novel approach to detect capsule types of the pneumococcus bacteria. 3-Minute Thesis Competition – Faculty of Health Final (QUT), August 2014, Kelvin Grove, Australia.

Rayner RE., Savill J., Hafner LM., Huygens F. The war against the pneumococcus. IHBI Inspires Postgraduate Student Conference, Annual Scientific meeting, November 2013, Brisbane, Australia.

Rayner RE., Savill J., Hafner LM., Huygens F. The population structure of Queensland invasive *Streptococcus pneumoniae* isolates in children: identified using a modified Multi-Locus Variable Number of Tandem Repeat Analysis BOX Typing. IHBI Inspires Postgraduate Student Conference, Annual Scientific meeting, November, 2012, Gold Coast, Australia.

Awards

Rayner RE., Savill J., Hafner LM., Huygens F. The war against the pneumococcus. 3-Minute Thesis Competition – Faculty of Health Final (QUT), August 2013, Kelvin Grove, Australia. (Runner-up finalist)

STATEMENT OF ORIGINAL AUTHORSHIP

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature:

Date: 28/07/2015

ACKNOWLEDGEMENTS

There are a number of people I would like to thank for assisting and enabling me to articulate from a Masters of Applied Science (Research) into a Doctor of Philosophy (PhD).

Firstly, I would like to thank my principal supervisor, Flavia Huygens, for her unfailing help and directing me on my studious path towards John Savill at the Queensland Health Forensic and Scientific Services. I am incredibly indebted for the opportunity to begin my study within a 'real-world' laboratory, and also the focussed push to begin practical work within my first two weeks of my Masters. I thank John for also being concerned about my welfare in the laboratory and assisting in my transfer back to IHBI, QUT, once I had articulated to my PhD.

I would also like to thank all of my three supervisors, Flavia, John and Louise, for their support in regards to writing papers, suggesting conferences (and pushing me to actually go), and encouraging and supporting any other activities associated to my PhD, university life and even my extracurricular lifestyle outside of university. I have found that the more things that I do (as long as I'm keeping a balanced schedule), the better I can apply my time and focus. So I appreciated their support in all the different areas and activities that I participated in.

I would like to thank the Public Health Microbiology, QHFSS members including Lawrence Ariotti, Megan Staples, Amy Jennison and Ricki Graham, for their assistance in the initial Masters research with culturing the bacteria and other assistance in experimental work. I would similarly like to thank the members of the Molecular Microbiology Pathogenesis Group (MMPG) - Sumeet, Irani, Imalka, Nadeesha, Asif, Leanne and James for their support and recommendations in our fortnightly meetings, and around the laboratory.

Finally I would like to thank the support of my family - Mum and Dad, my brothers Daniel, Luke and Josh, and Grandma, William, Nana, Marg, Lucy and Abigail. I also want to thank all my friends for supporting me in completing my PhD, showing interest in the subject I am researching and allowing me to de-stress away from my studies. Lastly and most importantly I would like to thank God and Jesus for being a constant and loving presence in my life.

CHAPTER 1: INTRODUCTION, AIMS, OBJECTIVES AND HYPOTHESES

1.1 Introduction

S. pneumoniae is a common but potentially deadly Gram-positive bacterium that causes high mortality and morbidity worldwide – over 800 000 children under five die from pneumococcal diseases each year globally (O'Brien *et al.,* 2009). In Australia, invasive pneumococcal disease (IPD) is still a major health risk especially in children under two years and in the Aboriginal population (Lehmann *et al.,* 2010).

An important feature of the pneumococcus is the polysaccharide capsule that surrounds the bacterium, protecting it from opsonophagocytosis (a mechanism of the host to clear out streptococcal infection via phagocytosis enhanced by an opsonin molecule) (Todar, 2012). Pneumococcal strains are often characterised by their capsule type, also called a serotype. Determining the serotype of pneumococci has traditionally been performed by a serotyping method e.g. Quellung reaction. The Quellung method uses polyclonal antibodies that bind to the capsule and produce a "swelling" reaction observed under a microscope (Statens Serum Institut, 2013). There are 98 published serotypes of *S. pneumoniae,* of which at least 23 serotypes are targeted by a vaccine, although potentially all 98 serotypes can cause disease (Bentley *et al.,* 2006; Bratcher *et al.,* 2010; Calix & Nahm, 2010; Calix *et al.,* 2012; Ko *et al.,* 2013; Oliver *et al.,* 2013; Park *et al.,* 2007; Park *et al.,* 2015).

To combat increasing incidences of IPD, a 7-valent pneumococcal conjugate vaccine (7vPCV; Prevenar®, Wyeth Pharmaceuticals) was introduced for all Australian children aged younger than two in 2005, targeting the seven most common serotypes (serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F) that caused IPD at that time (Johnson *et al.,* 2010). Since the introduction of the vaccine, IPD cases have decreased globally as expected, however IPD caused by non-vaccine (non-7vPCV) serotypes increased due to two phenomena: serotype replacement and capsule switching (Elberse *et al.,* 2011a; Elberse *et al.,* 2012; Guevara *et al.,* 2009; Hanage *et al.,* 2005; Hanna *et al.,* 2008; Hicks *et al.* 2007; Lehmann *et al.,* 2010; Marsh *et al.,* 2007; McChlery *et al.,* 2005; Nunes *et al.,* 2005; Reingold *et al.,* 2005; Singleton *et al.,* 2007; Tyrrell *et al.,* 2009; Whitney *et al.,* 2003).

Serotype replacement, the increase in non-vaccine targeted serotypes after the decrease in vaccine-targeted serotypes, has largely contributed to a shift in the pneumococcal population and increasing IPD (Weinberger *et al.,* 2011). Pneumococcal serotype 19A has been the most common non-7vPCV serotype causing IPD in the USA after the introduction of the 7vPCV (Pai *et al.,* 2005; Reingold *et al.,* 2005). The Queensland Health Pneumococcal Reference Laboratory has recorded that since 2007, there has been an increase in IPD cases caused by non-vaccine targeted serotypes from

113 (2000-2004) to 244 (2007-2011) in children under five years (raw numbers provided only - J. Savill, personal communication, 2012). From the previous observation, four serotypes included in an upgraded 13-valent pneumococcal conjugate vaccine (13vPCV; Pfizer) (targets the additional serotypes 1, 3, 7F and 19A) have contributed to over 140 IPD cases. Serotype 19A is mainly responsible for this increase and is the subject of another genotyping study at Queensland Health Forensic and Scientific Services (QHFSS). It is unknown whether there were any changes in pneumococcal surveillance, changes in guidelines for blood cultures, etc. that may have impacted this increase in non-vaccine targeted serotypes during this time period. However, the Commonwealth-funded Enhanced Invasive Pneumococcal Disease Surveillance Working Group was established in 2000 by the Communicable Diseases Network Australia to improve surveillance of IPD in Australia; all isolates of *S. pneumoniae* recovered from IPD are referred to reference laboratories, including QHFSS, for serotyping, except for the few that have been identified by PCR detection (Staples *et al.,* 2015). Serotype information of *S. pneumoniae* isolates sent to QHFSS was obtained for 94-99% of IPD cases during 2010 and 2012 for QLD, WA and NT (Staples *et al.,* 2015), indicating that the majority of isolates during this time period at least was identified.

Due to increases in 19A and other non-7vPCV serotypes, the 13vPCV was introduced in July 2011 in Australia, targeting an extra six serotypes as well as those targeted by the 7vPCV (Slaon-Gardner *et al.,* 2011). IPD caused by serotypes 8, 15B, 15C, 22F, 33F and 38 (non-13vPCV) have increased from 21 cases (2000-2004) to 46 cases (2007-2011) (QHFSS – Pneumococcal Reference Laboratory Database; raw numbers provided only). Studies have begun to emerge, examining what initial effects the 13vPCV is having on the pneumococcal population in respective countries or states. The outcomes of these studies are presented in Chapter 2 of this thesis. To date, no study has been published in Australia regarding the pneumococcal population structure after the introduction of the 13vPCV in July 2011.

The other phenomenon, capsule switching, is due to pneumococci switching capsule genes, which can result in a changed expressed capsule (Brueggemann *et al.,* 2007). Capsular switching is a natural phenomenon known to exist for a long time, clearly occurring before molecular typing methods existed (Avery *et al.,* 1944; Griffith, 1928; Nesin *et al.,* 1998). Multiple fragments of genetic material up to 44kb in length, including the capsule genes, have been reported to undergo recombination. In some settings there has been no evidence that capsule switching has increased following the introduction of PCVs (Simões *et al.,* 2011), however this ability to change the capsule

means that the vaccine no longer targets that serotype. This may lead to "vaccine escape" strains, already detected in the USA and Italy (Ansaldi *et al.,* 2011; Brueggemann *et al.,* 2007; Wyres *et al.,* 2013). The alarming ability to change the polysaccharide capsule via horizontal transformation renders the pneumococci effectively "immune" to the vaccine.

Therefore, bacterial fingerprinting methods, or more scientifically, genotyping methods are crucial to detect these capsule switches, whereas other traditional methods (e.g. serotyping) have limitations. Investigating the pneumococcal population has also been difficult in the past as no nationally agreed-on method has been accepted. It is important to monitor the changing epidemiology and population structure of *S. pneumoniae* in Australia and worldwide since the 13vPCV still only covers 13 serotypes out of at least 98 potentially invasive serotypes.

Currently, Multi-Locus Sequence Typing (MLST) is considered the "gold standard" for genotyping *S. pneumoniae*, superseding Pulse Field Gel Electrophoresis (PFGE) (Enright & Spratt, 1998; Malachowa *et al.,* 2005). MLST directly examines seven housekeeping genes and their single nucleotide polymorphisms (SNPs) that slowly change over time to characterise each pneumococcal strain. However, MLST is laborious, expensive and more suited to evolutionary studies rather than epidemiology studies (Hyytiä-Trees *et al.,* 2007). Another method, Multi-Locus Variable number of tandem repeat Analysis (MLVA) has been reported to be a more discriminatory genotyping method compared to MLST and PFGE (Elberse *et al.,* 2011a; Koeck *et al.,* 2005; Pichon *et al.* 2010; Van Cuyck *et al.,* 2012). MLVA detects Variable Number of Tandem Repeats (VNTRs) within the genome, and assigns a numeric profile to each pneumococcal isolate based on the length polymorphisms of each VNTR. There are at least four published MLVA methods for *S. pneumoniae,* each targeting different genetic regions in the pneumococcal genome (Elberse *et al.* 2011a; Koeck *et al.,* 2005; Rakov *et al.,* 2011; Van Cuyck *et al.,* 2012). Despite this, MLVA also has limitations particularly concerning the non-amplification of some loci, the lack of a large international database and no universally accepted method. Neither MLVA nor MLST has been used to determine the population structure of *S. pneumoniae* in Queensland after the introduction of the 13vPCV, and these will both be used in this study. By investigating the *S. pneumoniae* population structure post 13vPCV introduction, an understanding of the impact of the vaccine on *S. pneumoniae* genotypes will be possible.

In addition, serotyping methods are used to characterise the pneumococcus. However this technique is ambiguous, requires personal training, is expensive, cannot distinguish between all serotypes unless a combination of Factor serum are used, and

does not determine capsule switches without a genotyping method. As a consequence, many molecular serotyping methods utilising PCR have been developed such as long-template PCR (Bentley *et al.*, 2006), allele-specific PCR focused on specific serotypes (Bratcher *et al.,* 2011; Elberse *et al.,* 2011c; Mavroidi *et al.,* 2004), or multiplex-PCR (Ahn *et al.,* 2012; Pai *et al.,* 2006). These molecular methods directly examine the cassette of genes that code for each capsule. To date there is no universally accepted molecular serotyping method, and the Quellung reaction has remained as the universal serotyping technique.

This study will be focusing on determining the genotypic population of invasive *S. pneumoniae* prior to the introduction of the 13vPCV and a year after, using a modified MLVA4 technique and a novel capsular typing technique. From the fourteen published MLVA loci, the best combination (e.g. seven oligonucleotide loci) has been chosen for a modified MLVA4 method for genotyping invasive *S. pneumoniae* in Queensland. This modified MLVA4 aims to be more discriminatory, cheaper and faster than the "gold standard" MLST, as well as overcoming the problems in existing MLVA methods such as the failure to amplify certain loci. Determining the pneumococcal population structure using the modified MLVA4 method will allow us to determine whether any changes have occurred since the introduction of the 13vPCV. We expect that the new vaccine will apply selection pressure on the population structure and hence observable changes may be seen.

A novel capsular typing technique was also developed. A bioinformatics Minimum SNPs program, developed by researchers at Queensland University of Technology (QUT) was used. Minimum SNPs uses a binary marker system of T = present and A = absent of each capsule gene to create a "pseudoDNA sequence" (only contains A and T) for each of the known capsule cassettes with published sequences on the NCBI database (Robertson *et al.,* 2004; Stephens *et al.,* 2007). The program then analyses the pseudoDNA sequences and identified a minimum set of 'targets' (i.e. capsule genes) to determine each serotype. The capsule genes selected were amplified using real-time PCR to determine the serotype of invasive pneumococcal isolates. Combining the novel capsular typing method with MLVA4 genotyping allows the determination of capsule switching, which enhances information in epidemiology studies.

1.2 Hypotheses, Aims and Objectives

1.2.1 Hypotheses

There are four main hypotheses being investigated in this study:

1) The modified MLVA4 method for genotyping invasive *S. pneumoniae* will be more discriminatory, faster, cheaper and technically less-demanding than previously published MLVA methods and the 'gold standard' MLST.

2) The population structure of invasive *S. pneumoniae* isolated from Queensland children are clonally unrelated from post-7vPCV to post-13vPCV, meaning that the majority of the pneumococcal isolates will not share similar genetic profiles or form genetic complexes as demonstrated by MLVA4 and MLST typing. The rationale for this is that it is expected that the 13vPCV vaccine would provide selection pressure on the population structure.

3) Serotype replacement and capsule switching has occurred since the freely available 7vPCV in Australia 2005.

4) Capsule types and sub-types can be detected by a novel capsular typing method based on the MinimumSNPs program, and capsule switching can be verified by combining the novel capsular typing with modified MLVA4.

1.2.2 Aims and Objectives

Aim 1: To modify a Multi-Locus Variable number of tandem repeat Analysis (MLVA) procedure for genotyping invasive *S. pneumoniae*.

Objectives

a) To develop a faster, cheaper, highly discriminatory and technically less-demanding genotyping method for *S. pneumoniae* compared to the 'gold standard' MLST method, and improves on published MLVA methods by reducing the percentage of incomplete profiles due to non-amplified loci.

b) To ensure that a highly discriminatory MLVA4 method will provide a more accurate reflection of the *S. pneumoniae* population structure than what MLST currently provides.

Aim 2: To determine the population structure of invasive *S. pneumoniae* in Queensland isolated from children 15 years or younger post-7vPCV and one year post-13vPCV by MLVA, MLST and our modified MLVA4.

Objectives

a) To compare the 'gold standard' MLST genotyping method with published MLVA and the modified MLV4 method when used to represent the pneumococcal population structure.

b) To provide a current study and baseline of the Queensland invasive pneumococcal population, particularly post-13vPCV introduction in July 2011.

Aim 3: To determine whether the genetic population of non-vaccine (non-13vPCV) serotypes is changing in the Queensland pneumococcal population.

Objectives

a) To assess the initial and possible future impact of the newly introduced 13vPCV into Queensland.

b) To possibly predict which non-vaccine serotypes may become serotype replacement, indicating that a new vaccine may be required in the future. Public health strategies may also be influenced by these results.

Aim 4: To develop a novel capsular typing method based on a bioinformatics Minimum SNPs program to characterise the pneumococcal capsule cassette for each published pneumococcal serotype.

Objectives

a) To improve on current molecular capsule typing methods by using a Minimum SNPs program to select a minimum set of highly informative genetic targets to be utilised with real-time PCR.

b) To combine the novel capsular typing method with modified MLVA4 to determine capsule switching.

1.3 Significance

A number of significant points must be made regarding the four aims of this study. Firstly, the significance of developing and modifying existing MLVA genotyping methods will enable a potentially faster, cheaper and more discriminatory genotyping method for *S. pneumoniae*. A more discriminatory genotyping method will ensure a more accurate reflection of the *S. pneumoniae* population structure. MLST relies on stable housekeeping genes which are good for determining evolutionary and long term changes however it is not discriminatory enough to reveal changes in the short-term. Also, we want to improve on the published methods by minimising the percentage of

incomplete profiles due to non-amplified loci, thereby providing a more accurate reflection of the true pneumococcal population.

Secondly, since the introduction of 13vPCV in 2011, no studies have been published investigating invasive *S. pneumoniae* in Queensland. This study will examine any changes in the population structure by comparing post-7vPCV (2007-2010) and post-13vPCV (2011-2012). Results obtained from this study will contribute to MLST international database and MLVA database.

By examining in further detail the potential changes in the pneumococcal population structure, particularly the non-13vPCV population, we can determine whether any capsule switching or serotype replacement is emerging. Both these phenomena are important as they can result in increased IPD cases and vaccine escape strains that are not targeted by the current childhood vaccine. The significance of understanding these changes is that it may impact on further development of vaccines particularly in the choice of serotypes in future vaccines. Also public health strategies (e.g. advice to parents, child care centres, etc.) may be influenced by these results. Further study of the extra serotypes included in a new vaccine (15vPCV) that is in trial at the moment will enable us to determine whether this new vaccine could be a potential candidate for the next childhood pneumococcal vaccine or whether alternative vaccines need to be considered. It is important to understand this because it may impact the necessity of this new vaccine, such as an increase in serotype 22F and/or 33F targeted by the 15vPCV, or alternatively no increases or cases of IPD caused by these two extra serotypes.

Finally, the development of a novel capsule typing method, to be used in conjunction with MLVA genotyping to detect capsule switching, will provide a faster and cheaper serotyping method. Determining whether capsule switching has occurred as indicated by the results from previous aims of this study is important as it may indicate the emergence of a vaccine escape strain.

1.4 Account of scientific progress linking the chapters

This thesis is presented in the 'Thesis by monograph' style. It contains a literature review presented at two major milestones during the PhD candidature, an independent literature review article, and three Chapters written covering the four Aims of this study, with each chapter designed to be adapted into a publication. A final chapter summarises the major findings and future directions of this thesis.

Following the introductory chapter, **Chapter 2** is the literature review which focuses on an in-depth examination and discussion of *S. pneumoniae* and the associations of epidemiology with genotyping methods. Also stated at the beginning of Chapter 2 is a literature review article titled "Genotyping *Streptococcus pneumoniae*" that was published by Future Microbiology in April 2015 (Rayner *et al.,* 2015a). This review paper focuses on the current genotyping methods used for characterising *Streptococcus pneumoniae.*

Chapter 3 contains the general methods used and developed for this thesis. Each chapter hereafter still contain a brief methodology section as well as any specific methodology for the aim.

Since there was no nationally accepted genotyping method of *S. pneumoniae* in Australia, examination of the current methods available was performed as the first contributions to this PhD. **Chapter 4** examines the current genotyping methods and the modification of MLVA to minimise its limitations (relates to Aim 1). Stated at the beginning of Chapter 4 is a scientific article titled "Modified MLVA for genotyping invasive *Streptococcus pneumoniae* in Queensland children" that was published by PLoS One in April 2015 (Rayner *et al.,* 2015b). This research was also presented as an abstract (poster form) at the IHBI Inspires QUT conference at the Gold Coast (November 2012).

By using our modified MLVA4 (Chapter 4) to determine the population structure of *S. pneumoniae* in Queensland, we continued to research whether the Queensland *S. pneumoniae* population structure was changing from 2007 to 2012, particularly since a new childhood vaccine, 13vPCV, was recently introduced in 2011. **Chapter 5** examines these changes in the pneumococcal population (relates to Aim 2 and Aim 3). This abstract was presented in poster form at the Australian Society for Microbiology conference held in Adelaide (July 2013) and at the Microbiology at QUT and Beyond Scientific Meeting in Brisbane (October 2014). An abstract was presented in oral form at the IHBI Inspires QUT conference in Brisbane (November 2013) and the 3-Minute Thesis Competition Faculty of Health Final QUT in Brisbane (August 2013).

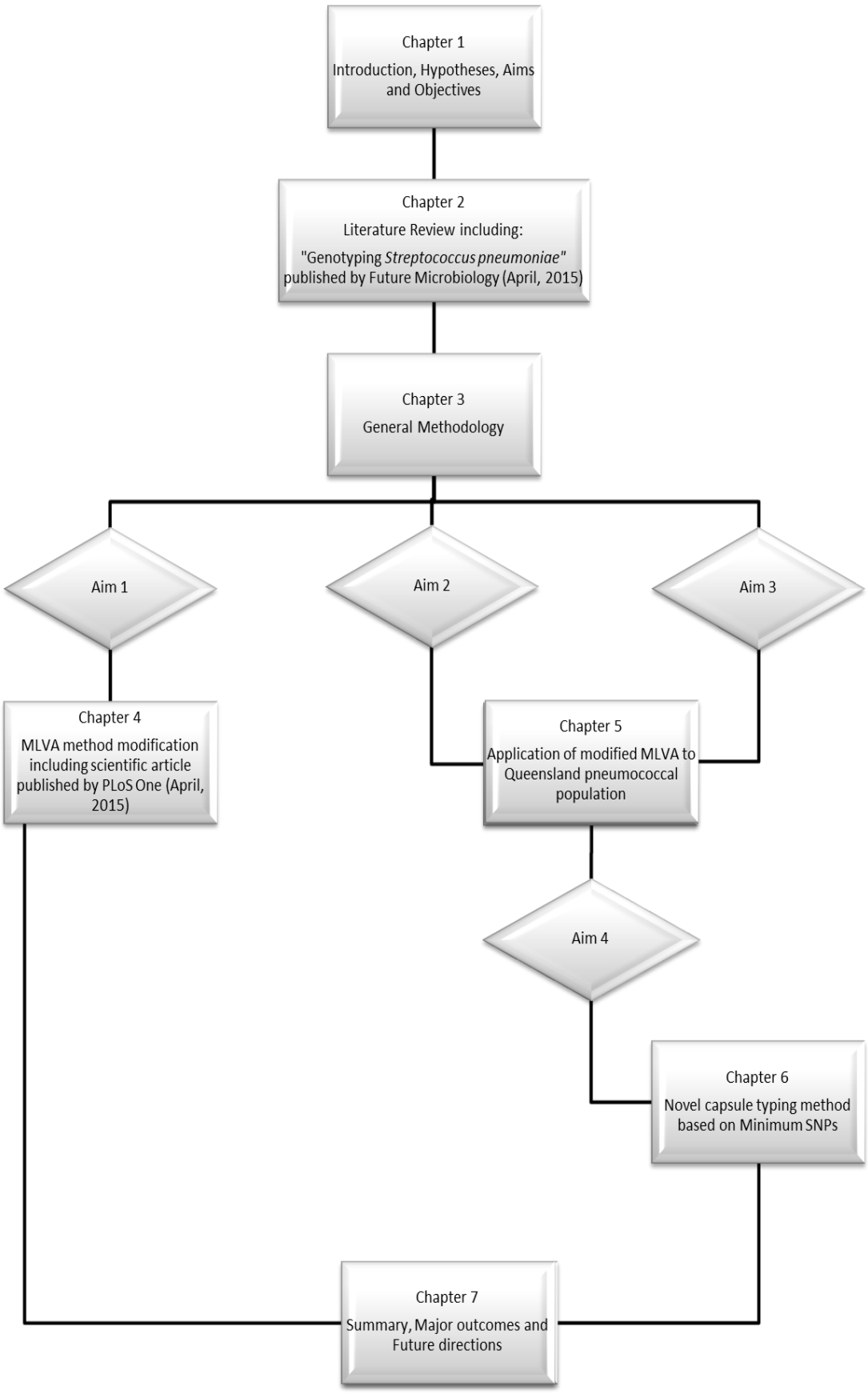
From our findings and published research, since capsule switching has been noticed in the *S. pneumoniae* population structure, we wanted to improve capsule typing methods by focusing on molecular typing but take a novel approach by using the Minimum SNPs program. **Chapter 6** examines the research and development of this novel capsular typing method which has been applied to the Queensland pneumococcal isolates. This novel capsule typing method was combined with the MLVA4 genotyping

method to determine whether capsule switching has occurred in Queensland (relates to Aim 4). This abstract was presented at the XIX Lancefield International Symposium of Streptococci and Streptococcal Diseases in Buenos Aires, Argentina (November 2014 - oral presentation), the QUT IHBI Inspires Postgraduate Student Conference at the Gold Coast, Australia (November 2014 - oral presentation), the Australian Society for Microbiology state meeting Microbiology in Maleny at Maleny, Australia (November 2014 - oral presentation), the Australian Society for Microbiology Annual Conference at Canberra, Australia (July 2015 – poster presentation) and the 3-Minute Thesis Competition Faculty of Health Final QUT in Brisbane (August 2014).

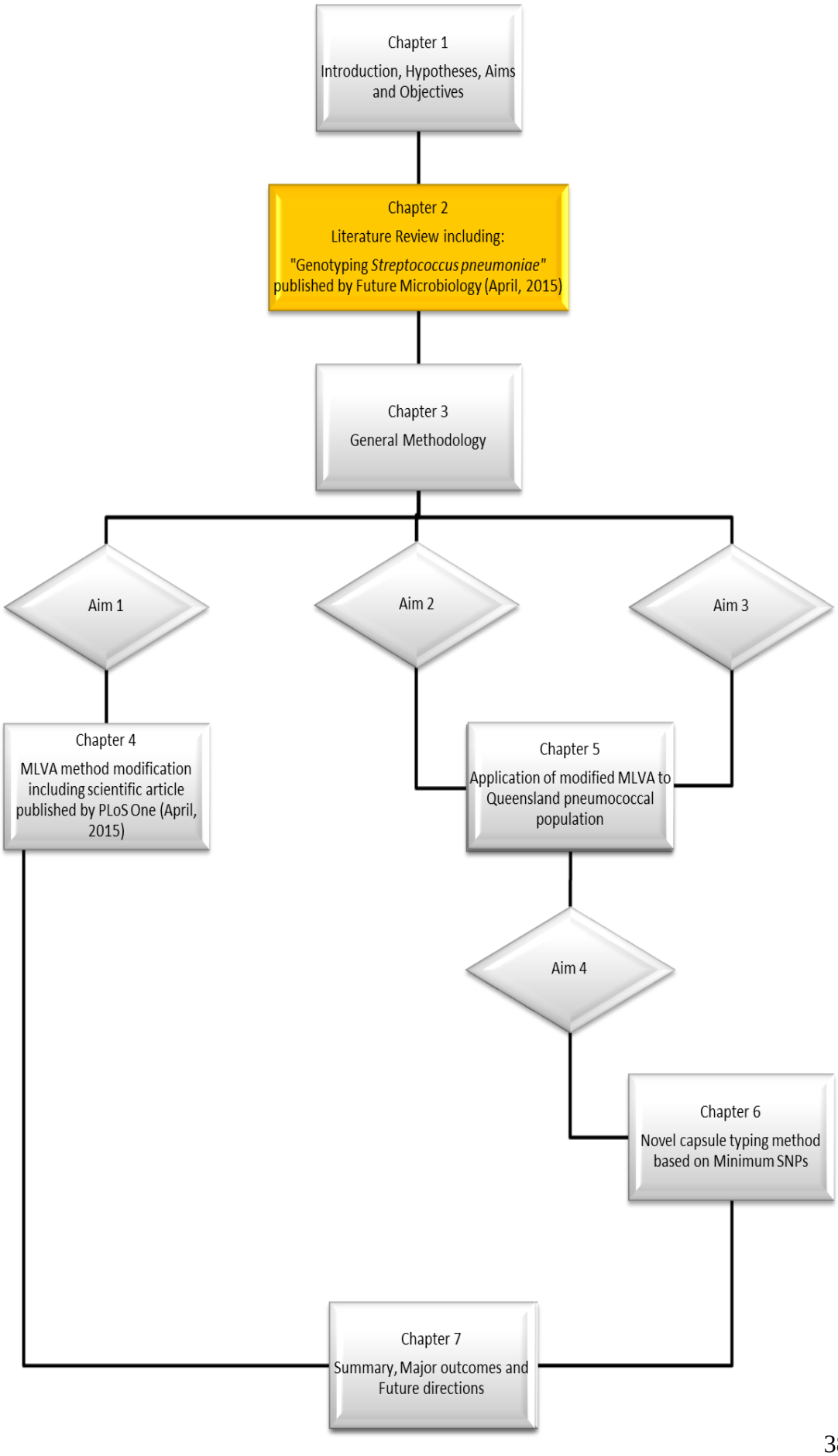
Chapter 7 summarises the over-arching discussion points from the literature review and three following chapters, the major findings and future directions.

The relationship between the various chapters/papers is shown in Figure 1.

Figure 1: Outline of chapters and aims linking the PhD work.



CHAPTER 2: LITERATURE REVIEW


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

2.1 Introduction

A review paper entitled: "Genotyping *Streptococcus pneumoniae*", published in Future Microbiology (April, 2015), is available (*Appendix B8: Published manuscript "Genotyping Streptococcus pneumoniae" to Future Microbiology (April 2015)*) which outlines the current *S. pneumoniae* genotyping methods and highlights the importance of further development of a robust and universal genotyping method (Rayner *et al.,* 2015a). This review is designed to be part of the literature review. The following literature review is updated from what was written for the Stage 2 milestone and confirmation/articulation of the PhD. It contains information regarding the bacterium *Streptococcus pneumoniae*, its prevalence in Australia, and the trends observed in epidemiology studies worldwide and nationwide.

2.2 *Streptococcus pneumoniae*

Streptococcus pneumoniae is a medically important bacterium as it can cause numerous diseases such as invasive pneumococcal disease (IPD) (including bacteraemia and meningitis), pneumonia, and otitis media (Hermans *et al.,* 1995; Koeck *et al.,* 2005; Todar *et al.,* 2012; Yaro *et al.,* 2006). This Gram-positive diplococcus bacterium normally inhabits the human upper respiratory tract (URT) without triggering any symptoms. However *S. pneumoniae* is associated with high morbidity (estimated 14.5 million episodes of serious pneumococcal disease in 2000), high mortality (11% of deaths in children aged younger than five, globally) and high economic burden especially in underdeveloped countries (O'Brien *et al.,* 2009).

An important feature of this bacterium is the protective polysaccharide capsule that forms the outer layer surrounding the bacterial cell. The pneumococcal capsule is also referred to as a serotype, a name given from the serological typing method to distinguish each capsule type. There are at least 98 known serotypes of *S. pneumoniae*, and a number of serotypes are associated with high mortality and morbidity (Bentley *et al.,* 2006; Bratcher *et al.,* 2010; Calix & Nahm, 2010; Calix *et al.,* 2012; Ko *et al.,* 2013; Oliver *et al.,* 2013; Park *et al.,* 2007; Park *et al.,* 2015). The pneumococcal capsule is an important virulence factor as it can provide resistance to opsonophagocytosis (Bentley *et al.,* 2006; Elberse *et al.,* 2011c; Kelly *et al.,* 1994; Sabharwal *et al.,* 2014; Schouls & Van der Heide, 2012), although Jefferies *et al.* (2004) theorised that clonal properties may also be important in invasiveness. It is believed that the large number of different capsular serotypes is due to a mechanism to evade the human immune response (Bentley *et al.,* 2006).

Beside from resistance to opsonophagocytosis, some capsular types are more invasive than others as they have been shown to have increased binding to a specific host protein inhibitor factor H (FH) (Hyams *et al.,* 2013; Scott *et al.,* 1996). For example, serotype 19A is particularly virulent and causes the majority of pneumococcal disease in USA, especially one particular clonal cluster CC199 (Pai *et al.,* 2005). Increases in serotype 19A in USA were also accompanied by significant increases in penicillin non-susceptibility and multi-drug resistance (Pai *et al.,* 2005). Since serotype 19A was not targeted by early vaccines, there has been a huge effort involved in monitoring this serotype. However other countries have significant increases in other serotypes, not serotype 19A.

As well as this, some serotypes are more commonly isolated from different clinical sites, for example pneumococcal serogroup 1 and 14 are commonly isolated from blood while others are commonly isolated from middle ear fluid (serogroup 3, 19 and 23) (Hausdorff *et al.,* 2000b; Scott *et al.,* 1996). A study reported that there may be two classes of *S. pneumoniae* – major invasive serotypes such as serotypes 1, 4, 7F and 9V that are highly clonal and found to be geographically confined, and common carriers such as serotypes 6A, 6B, 14 and 19F (Sá-Leão *et al.,* 2011; Sandgren *et al.,* 2004). Two other key studies by Hausdorff *et al.,* (2000a; 2000b) were fundamental in showing the serotypes most frequently associated with invasive disease and determining vaccine formulations. These studies indicated that at the time, the serotypes targeted by the 7vPCV caused 70%-88% of IPD in young children in the USA and Canada, Oceania, Africa and Europe (Hausdorff *et al.,* 2000a). They also pointed out that several serogroups that were not targeted by the 7vPCV, 9vPCV or 11vPCV were significant causes of disease in older children and adults (Hausdorrf *et al.,* 2000a). By understanding the prevalence and virulence of these major worldwide serotypes, this has led to further development of vaccines to target them.

In addition, there are studies that clearly show the importance of the genotype in invasiveness. Irrespective of the serotype's overall invasive potential, genetic clones displayed heterogeneous behaviour in invasiveness (Sá-Leão *et al.,* 2011). This highlights the importance that both the serotype and the genetic clonal type determine invasiveness. Even biofilm formation varied between isolates with the same serotype and same genotype, indicating that different clinical sites (e.g. blood or ear) have an impact on colonisation and invasiveness (Trapetti *et al.,* 2013).

2.3 Invasive pneumococcal disease (IPD)

The World Health Organisation (WHO) estimated that 1.6 million people died from pneumococcal diseases in 2005, and it is estimated that more than 800 000 children under five die each year globally from pneumococcal disease (O'Brien *et al.,* 2009). IPD encompasses a range of diseases caused by *S. pneumoniae*, including meningitis and septicaemia. Pneumonia is the most common pneumococcal disease, although is not termed as "invasive" in the sense of being isolated from a sterile body site (Centers for Disease Control Prevention, 2015). The most common pneumococcal diseases observed are pneumonia and meningitis in children, and pneumonia in adults (O'Brien *et al.,* 2009). The symptoms of IPD vary depending on the specific disease and can be non-specific in children, even including sporadic fever (Randle *et al.,* 2011).

In Queensland, Australia, all cases of IPD are referred to the Queensland Health Pneumococcal Reference Laboratory, Brisbane, for serological typing. The notification rate of IPD in Queensland in 2014 was 4.9 per 100 000, a decline from 7.6 per 100 000 in 2011 (NNDSS, 2015). Australian Aborigines still maintain the highest rates of IPD in the worldwide.

For clinicians to make a diagnosis of IPD, the pneumococcus needs to be isolated from normally sterile body sites, for example blood, cerebrospinal fluid (CSF), joint fluid, pericardial fluid, pleural fluid and certain tissues (pneumococci can colonise the nasopharyngeal tissue without causing disease) (Centers for Disease Control Prevention, 2015). Diagnostic of IPD is obtained by culture or detecting bacterial cell surface antigens (Randle *et al.,* 2011). Depending on the severity of the disease and the type of disease, treatment will vary. Generally, antibiotic treatment is provided for most pneumococcal diseases. Unfortunately, a number of pneumococcal strains have developed resistance to certain antibiotics, as further discussed in section 2.4.

2.4 Introduction of Public Health Preventative Measures to control *Streptococcus pneumoniae*

Due to high rates of mortality and morbidity, a number of preventative measures have been implemented worldwide to combat pneumococcal diseases. Antibiotic therapy was one of the first public health preventative measures against *S. pneumoniae* after the Second World War (WWII). Unfortunately this also led to the first description of a penicillin-resistant *S. pneumoniae* in Australia in 1967 (Hansman & Bullen, 1967). Since then, penicillin-resistant strains have spread across the globe. Multiple-drug

resistant (MDR) pneumococcal strains, first reported in 1977 in Johannesburg, South Africa, have also populated the world (Jacobs *et al.,* 1978). The first description of a MDR *S. pneumoniae* (serotype 6B) in Australia was in Far North Queensland in February 1995 (Hanna *et al.,* 1997).

One mechanism to acquire antimicrobial resistance is the horizontal gene transfer of resistant genes on chromosomally located transposons, not plasmid mediated like most other bacterial species (Cornick & Bentley, 2012; Hall *et al.,* 1998). Horizontal transformation or recombination can rapidly introduce novel phenotypes, and is thus a major driving force behind the evolution of *S. pneumoniae,* particularly in response to stressful conditions such as antimicrobials (Cornick & Bentley, 2012). Other mechanisms include hyper-mutation and hetero-resistance (Cornick & Bentley, 2012). Hyper-mutation phenotype can lead to an increased rate of DNA mutations, for example, modification of penicillin binding proteins (PBPs) reduces affinity to penicillin drug (Cornick & Bentley, 2012). Hetero-resistance is a function not fully understood in pneumococci but has been studied in Staphylococci, where a subpopulation has the ability to tolerate higher antimicrobial conditions (Cornick & Bentley, 2012).

The Pneumococcal Molecular Epidemiology Network (PMEN) was established in 1997 to provide a nomenclature to the most important pneumococcal sequence types that were MDR (Klugman *et al.,* 1990; McGee *et al.,* 2001; PMEN, 2015). There is a criterion for the assignation of PMEN to a *S. pneumoniae* strain (Figure 2).

Figure 2: PMEN criteria for *S. pneumoniae* (PMEN, 2015).

- Clone should have wide geographic distribution (≥2 continents)
- Clone resistant to one or more antibiotics that are in wide clinical use; or a global susceptible clone known to be important in disease
- Data on clone published
- New clones proposed at annual PMEN meeting
- Representative isolate made available for analyses and confirmation
- Available to deposit in ATCC collection of clones
- Available through Streptococcus Reference Laboratory in Germany and Emory University in USA

Since the discovery of MDR *S. pneumoniae*, vaccines have been developed that target the polysaccharide capsule (Hanna *et al.,* 1997). Interestingly, the earliest vaccine against *S. pneumoniae* was developed more than 100 years ago using whole pneumococcal cell to vaccinate South African miners in 1911 (Wright *et al.*, 1914). Several decades later a four-valent pneumococcal polysaccharide vaccine (PPV) was trialled in military recruits before developing a six-valent PPV to vaccinate the military

from 1946 to 1948 (MacLeod *et al.*, 1945; Moberley *et al.,* 2013). Trials continued to vaccinate South African gold miners with a 13-valent PPV (Austrian, 1977) and Papua New Guinean highlanders with a 14-valent PPV (Riley *et al.*, 1977), which was licensed in the USA the same year.

In more recent times, a 23-valent PPV (23PPV; Pneumovax®23, Merck & Co. Inc.) was developed in 1983 to target 23 serotypes (Table 1) (Moberley *et al.*, 2013; Robbins & Schneerson, 1983; Roche *et al.,* 2003; Whitney *et al.,* 2003;). However, the 23PPV was poorly immunogenic in children, which was problematic since the majority of pneumococcal diseases occurred in children (Roche *et al.,* 2003; Swiatlo & Ware, 2003).

Table 1: The *S. pneumoniae* serotypes that are targeted by the available polysaccharide conjugate vaccines (PCV) or pneumococcal polysaccharide vaccine (PPV).

Pneumococcal vaccine	Targeted *S. pneumoniae* serotypes	Production company	Commercially available in AUS
7vPCV; Prevenar®	4, 6B, 9V, 14, 18C, 19F, 23F	Wyeth Pharmaceuticals Inc.	No
9vPCV	7vPCV serotypes + 1, 5	Wyeth Pharmaceuticals Inc.	No
10vPCV; Synflorix®	7vPCV serotypes + 1, 5, 7F	GlaxoSmithKline Australia	Yes
11vPCV	7vPCV serotypes + 1, 3, 5, 7F	Sanofi Pasteur	No
13vPCV; Prevenar 13	7vPCV serotypes + 1, 3, 5, 6A, 7F, 19A	Pfizer Australia Pty Ltd	Yes
15vPCV (in trial)	13vPCV serotypes + 22F, 33F	Pfizer Australia Pty Ltd	No
23PPV; Pneumovax® 23	1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, 33F	Merck & Co. Inc.	Yes

As a result, the 7vPCV (Prevenar®, Wyeth Pharmaceuticals Inc.) was developed. It was licensed for use in Australia in 2001 after a USA trial demonstrated vaccine efficacy of 93.9% in 2-15 month old children (Hills *et al.,* 2002). A conjugate vaccine consists of an immunogenic carrier protein linked to polysaccharide antigens, which elicits an immune memory response through antigen-specific memory B-cell production (Duggan, 2010). The 7vPCV couples a mutated diphtheria toxoid to a protein carrier CMR_{197} (Centers for Disease Control Prevention, 2010). The difference between the PPV and PCV is that the PCV contains the serotype antigens conjugated to protein so that

children have a higher response when forming antibodies, whereas the PPV only has antigens which by themselves failed to induce high levels of antibodies in children (Swiatlo & Ware, 2003). Unfortunately, due to the nature of the vaccine, only seven serotypes were targeted (Table 1) (Roche *et al.,* 2003; Whitney *et al.,* 2003). Countries with different dominating serotypes than those targeted by the 7vPCV and increasing IPD caused by non-7vPCV serotypes prompted the development of the 9vPCV, 10vPCV, 11vPCV and 13vPCV, as further discussed below.

The 9vPCV (Wyeth Pharmaceuticals Inc.) targeted additional serotypes 1 and 5 which are prominent in the UK and USA, but not Australia, however the vaccine itself has never been commercialised (Goldblatt *et al.,* 2006). The 10vPCV (Synflorix®, GlaxoSmithKline) includes all the 7vPCV serotypes as well as serotypes 1, 5 and 7F, however in Australia the vaccine was only introduced to Northern Territory vaccination programs since serotype 5 is not common elsewhere in Australia (Hanna *et al.,* 2010). Eventually, the 13vPCV (Prevenar 13®, Pfizer Australia Pty Ltd.) was introduced in the USA in February 2010 and Australia in 2011 (Duggan, 2010; Selvey, 2011). The 13vPCV contains an additional six serotypes as well as those targeted by the 7vPCV. Serotypes are again conjugated to a non-toxic diphtheria protein with cross-reactive material, CRM_{197} (Table 1) (Centers for Disease Control Prevention, 2015; Duggan, 2010; Hanna *et al.,* 2010; Rakov *et al.,* 2011;).

The 13vPCV elicited strong immunological responses against all 13 serotypes (Duggan, 2010). A polysaccharide conjugate vaccine has been added to routine infant immunisation schedules of 112 (58%) of 194 WHO member states as of November 2014 with the help of funding from Global Alliance for Vaccines and Immunisation (GAVI) Alliance (Table 2) (Murray *et al.,* 2014). An additional benefit of the vaccine is that even though children are vaccinated at 2, 4 and 12 months of age (or a variation in other countries), by the time they reach pre-school age the majority of children have been shown to maintain protective levels of serotype-specific antibodies (Trück *et al.,* 2014).

Table 2: The year that countries have introduced a Polysaccharide Conjugate Vaccine (PCV) into their National Immunisation Program, or have made the vaccine commercially available only.

BFA (Burkina Faso); CAF (Central African Republic); COD (Congo, The Democratic); CZE (Czech Republic); DOM (Dominican Republic); GNB (Guinea-Bissau); LBY (Libyan Arab Jamahiriya); PNG (Papua New Guinea); STP (Sao Tome and Principe); TTO (Trinidad and Tobago); ARE (United Arab Emirates); UK (United Kingdom); USA (United States of America)

Country/state	7vPC	13vPCV	Country/state	7vPCV	13vPCV	Country/state	7vPCV	13vPCV
Afghanistan	-	2013	France	2006	2010	Mozambique	-	2013
Alaska	2001	2010	Gambia	2009	-	Netherlands	2006	-
Albania	-	2011	Georgia	-	2013	New Zealand	2008	-
Andorra	2007	-	Germany	2006	-	Nicaragua	2010	-
Angola	-	2013	Ghana	-	2012	Niger	-	2013
Argentina	-	2012	Greece	2005	-	Nigeria	-	2013
Armenia	-	2013	Greenland	2007	-	Niue	2009	-
Australia	2005	2011	Guatemala	-	2012	Norway	2006	-
Austria	2002	-	GNB	-	2013	Oman	2008	-
Azerbaijan	-	2013	Guyana	-	2011	Pakistan	-	2012
Bahamas	2009	-	Haiti	-	2013	Palau	2008	-
Bahrain	2008	-	Honduras	-	2011	Panama	2010	-
Bangladesh	-	2013	Hungary	2009	-	PNG	-	2013
Barbados	2009	-	Iceland	2011	-	Paraguay	-	2012
Belgium	2006	-	Ireland	2008	-	Peru	2009	
Belize			Israel	2009	-	Philippines	-	2013
Benin	-	2011	Italy	2005	-	Portugal	2001	-
Bolivia	-	2014	Jamaica	-	2010	Qatar	2005	-
Botswana	-	2012	Japan	2011	-	Rwanda	2009	-
Brazil	2010	-	Jordan	2011	-	STP	-	2012
Bulgaria	2010	-	Kazakhstan	2011	-	Saudi Arabia	2009	-
BFA	-	2013	Kenya	-	2011	Senegal	-	2013
Burundi	-	2011	Kiribati	-	2013	Sierra Leone	-	2011
Cambodia	-	Apply	Korea, Rep.	2004	-	Singapore	2009	-
Cameroon	-	2011	Kuwait	2007	-	Slovakia	2009	-
Canada	2002	-	Lao PDR	-	2013	Slovenia	2005	-
CAF		2011	Latvia	2010	-	South Africa	2009	-
Chile	2009	2011	Lebanon	2010	-	Spain	2003	-
Colombia	-	2011	Lesotho	-	2013	Sudan	-	2013
Comoros	-	Apply	Liberia	-	2014	Sweden	2009	-
COD	-	2011	LBY	-	2013	Switzerland	2006	-
Congo	-	2012	Lithuania	-	2011	Tanzania	-	2012
Costa Rica	2008	-	Luxembourg	2005	-	Togo	-	2013
Cyprus	2007	-	Madagascar	-	2011	TTO	2009	-
CZE	2005	2010	Malawi	-	2011	Turkey	2008	-
Denmark	2007	-	Mali	-	2011	Uganda	-	2013
Djibouti	-	2012	Malta	2007	-	ARE	2007	-

Country/state	7vPC	13vPCV	Country/state	7vPCV	13vPCV	Country/state	7vPCV	13vPCV
Dominica	-	2010	Marshall Is.	2009	-	UK	2006	2010
DOM	2009	-	Mauritania	-	2013	USA	2000	2010
Ecuador	-	2010	Mexico	2009	-	Uruguay	2008	2010
El Salvador	-	2010	Micronesia	2008	-	Yemen	-	2011
Ethiopia	-	2011	Moldova, R.	-	2013	Zambia	-	2013
Fiji	-	2012	Mongolia	-	2012	Zimbabwe	-	2012
Finland	-	2010	Morocco	2010	-			

2.5 Impact of the Pneumococcal Conjugate Vaccines Globally

Worldwide, including Australia, epidemiology studies have been conducted to assess the local impact of these polysaccharide vaccines on the pneumococcal population structure. Since the introduction of the 7vPCV globally, there were reports of an overall decrease in IPD in children, which was a positive outcome of vaccine administration (Poehling *et al.,* 2006; Reingold *et al.,* 2005; Singleton *et al.,* 2007; Whitney *et al.,* 2003). For instance, in Native Alaskan children under two years there was an overall 67% decrease in IPD between 2002 to 2004 after the introduction of the 7vPCV in January 2001 (96% were 7vPCV serotypes) (Singleton *et al.,* 2007). Similarly, in the USA there was a 78% drop in 7vPCV serotypes in children under two years by 2003 (licensed in 2000 for under five years) (Reingold *et al.,* 2005; Whitney *et al.,* 2003).

This decrease in IPD in children has either been a direct result of the introduction of the vaccine, or an indirect effect such as herd immunity or herd protection (Hanna *et al.,* 2008; Isaacman *et al.,* 2007). Reingold *et al.* (2005) theorised herd immunity when 69% of IPD in their study was prevented through indirect effects of the vaccine, possibly because vaccinated children are less likely to transmit colonising pneumococci to other members of the family. Similar results of herd immunity have been reported elsewhere (Rodgers & Klugman, 2011). Since then, there has been a significant impact on IPD and carriage by herd immunity (Davis *et al.,* 2013).

Interestingly, Hicks *et al.* (2007) stated that the 7vPCV also has residual effects; non-vaccine serotypes 6A and 12F decreased significantly after the introduction of the 7vPCV despite not being included in the vaccine. Serotype 6A has been shown to decrease due to cross-protection against 6B, however since the detection of serotype 6C, it has been difficult to determine the impact of the 7vPCV on serogroup 6 (Park *et al.,* 2007). On the other hand, it is thought that the decrease in 12F is due to natural fluctuations over time. Hence any increases in 12F may not necessarily be due to serotype replacement. Serotype 12F is often reported in outbreaks of IPD rather than

gradual replacement of vaccine-serotypes in the population and is considered an 'epidemic' serotype (Gratten *et al.,* 1995; Jorgensen *et al.,* 2005).

Following the introduction of the new 13vPCV in the USA, France and the UK in 2010, reports are emerging that investigate the effects of this expanded vaccine on the *S. pneumoniae* population. In France there was a noticeable drop in non-7vPCV serotypes (19A, 6C and 7F) in carriers with at least one dose of 13vPCV compared to children only vaccinated with the 7vPCV (Cohen *et al.,* 2012). Two reports have observed decreases in 13vPCV serotypes in Alaska, one observing a significant decline of 13vPCV *S. pneumoniae* in nasopharyngeal colonisation of Alaskan rural children (25% to 5%) (Gounder *et al.,* 2014; Singleton *et al.,* 2013). A decline in 13vPCV was also observed in Madrid, Spain, from 2010-2011 compared to 2007-2010 (Picazo *et al.,* 2013), and in Germany where an overall decline in 13vPCV serotypes in 2010 to 2011 was observed in children despite an increase in serotype 19A (Van der Linden *et al.,* 2013).

The 13vPCV has also been shown to hold residual effects similar to the 7vPCV. Cooper *et al.* (2011) and Miller *et al.* (2011) determined that serotype 6A, included in the 13vPCV, provides cross-protection for serotype 6C, as well as cross-protection of serotypes 7A by 7F. A noticeable drop in serotypes 7F and 19A was observed in the UK after the introduction of 13vPCV in 2010 (Miller *et al.*, 2011). In the USA, after the 13vPCV was introduced in 2010, there has been a 53% decline in IPD in children aged under two years compared to the calculated average rate of IPD in 2007 to 2009 (Kaplan *et al.,* 2013). The decline in IPD was attributed to decreases in serotype 19A (58%), serotype 7F (54%) and serotype 3 (68%), all targeted by the 13vPCV (Kaplan *et al.,* 2013). Currently, there are no reports from Australia regarding the impact of the 13vPCV. Despite these promising results from other countries, further studies around the globe are required to determine the short-term and long-term effects of the 13vPCV.

2.6 Impact of the Pneumococcal Conjugate Vaccines in Australia and Queensland

Australia has kept up-to-date with the introduction of the polysaccharide conjugate vaccines, particularly since Australian Aborigines are 6.7 times more likely to have IPD than non-Aboriginals in 2010, demonstrating that Australian Aborigines still suffer from IPD more than non-Indigenous citizens (Giele *et al.,* 2007; Lehmann *et al.,*

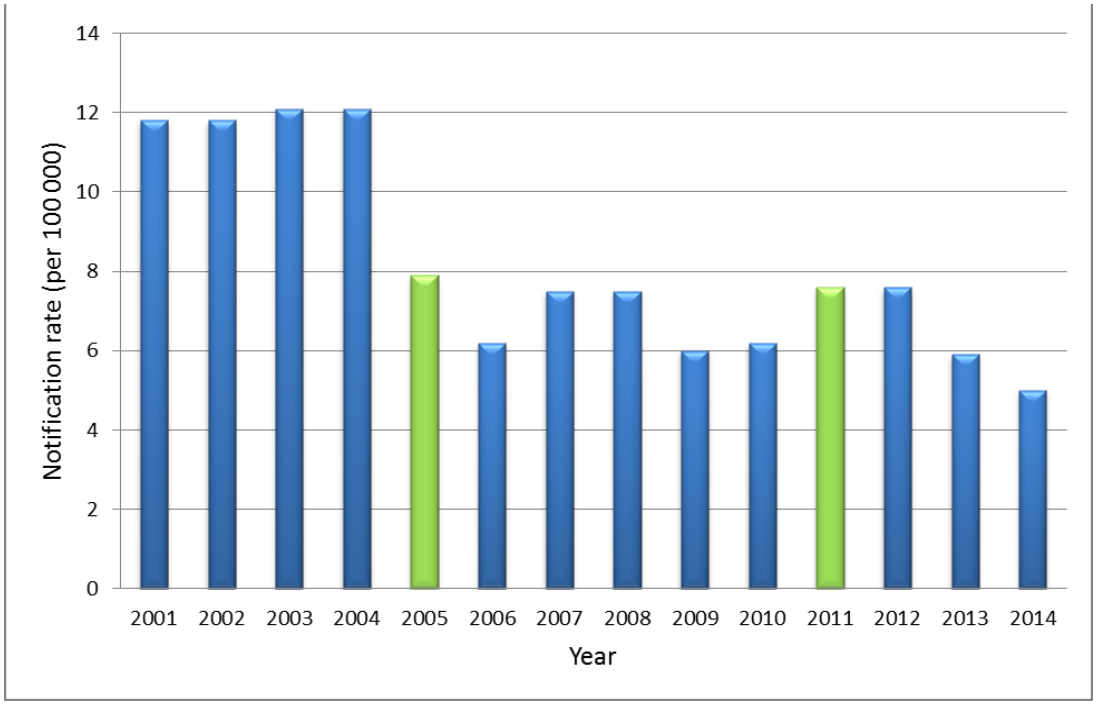
2010). The 23PPV was introduced for at-risk Indigenous adults in 1986 (Hanna *et al.,* 1997). However, since the 23PPV is not as effective in children, the 7vPCV was introduced for Australian Aboriginal children younger than two years in 2001 and became freely available for all children in 2005. The 7vPCV was given at 2, 4 and 6 months of age, followed by a booster dose of 23PPV (Pneumovax23®) only for Aboriginal and Torres Strait Islander children living in NT, SA, QLD and WA, and children with specified underlying medical conditions, administered between 18-24months (NCIRS, 2015). Invasive diseases caused by the 7vPCV serotypes decreased by 78% in children younger than two between 2002 and 2006, mirroring statistics observed elsewhere in the world (Lehmann *et al.,* 2010, Roche *et al.,* 2008).

Decreases of IPD in 2-14 year olds by 2005 were also observed, indicating herd protection since these older children did not have access to the 7vPCV (Roche *et al.,* 2007). There was a more pronounced decrease of IPD cases in non-Indigenous children than Indigenous children, most likely because there was a wider range of serotypes in the Indigenous population and the vaccine only targeted a few of these serotypes (Fagan *et al.,* 2001; Roche *et al.,* 2007; Williams *et al.,* 2011).

Eventually, the National Notifiable Diseases Surveillance System (NNDSS) began recording the number of IPD cases in Australia from 2001. By 2012, the NNDSS recorded 7.6 per 100 000 notification rate of IPD in Queensland, an increase from 6.0 per 100 000 notifications in 2009, possibly due to increasing IPD caused by non-7vPCV serotypes (Figure 3) (NNDSS, 2015). Roche *et al.* (2007) predicted that rates of IPD cases under two years may not fall much further with only the 7vPCV. By 2007, the 7vPCV only covered 55% of serotypes under two year old Indigenous children but 86% in non-Indigenous children, highlighting the difference in vaccine effectiveness. A decline in pneumococcal disease can be seen after the introduction of the 7vPCV in 2005 and possibly after the 13vPCV in 2011 (Figure 3).

Figure 3: Invasive pneumococcal disease (IPD) notification rates in Queensland from 2001 to 2014 as recorded by the National Notifiable Diseases Surveillance System.

Green bars represent the introduction of 7vPCV in 2005 and 13vPCV in 2011.



Similarly, epidemiology studies observed a 98% decline in IPD caused by 7vPCV serotypes in Australian children from 2002-2004 compared to 2007-2009, however there was a slight increase in 13vPCV serotypes (Johnson *et al.,* 2012). Barry *et al.* (2012) reported a 168% increase in non-7vPCV serotypes in children aged under five in Australia in 2008 compared to 2002, with a fourfold increase due to serotype 19A. This was not promising news.

Therefore, the 13vPCV was licensed for use in Australia and was introduced into the Prevenar13® Immunisation Program for 2, 4 and 6 month old Australian children in October 2011 (Selvey *et al.,* 2011). A fourth dose of 13vPCV could be administered as a booster for Aboriginal and Torres Strait Islander children at age 12-18months living in NT, SA, QLD and WA, replacing the 23PPV booster shot (NCIRS, 2015). By 2013, the NNDSS notification rates of IPD in Queensland had appeared to drop to 5.9 per 100 000 notifications, the lowest rates ever since records were kept from 2001 (Figure 3). Even in 2014, decreasing rates of IPD have been documented, with notifications of 5.0 per 100 000 by the end of December being recorded (NNDSS, 2015). During May 2014, the latest change to the immunisation schedule was the extension of the 13vPCV registered use in children from 6 weeks of age (NCIRS 2015). To date, there have been no other

reports on the impact of the 13vPCV on the pneumococcal population; however this may be due to the recent introduction of this vaccine.

2.7 Future Pneumococcal Vaccines

Despite the significant decreases in IPD due to 7vPCV, Hicks *et al.* (2007) suggested that multivalent vaccines should be expanded further to include the common replacement serotypes, particularly those prevalent in third world countries. Therefore additional multivalent vaccines have been produced including 9vPCV, 10vPCV, 11vPCV and 13vPCV. Additionally, pre-clinical trials of a 15-valent pneumococcal conjugate vaccine (15vPCV) on infant rhesus monkeys have been carried out (Skinner *et al.,* 2011). The serotypes included in the 15vPCV are those in the 13vPCV plus serotypes 22F and 33F, the "two most prevalent emerging serotypes" in the USA (Skinner *et al.,* 2011). The prevalence of different serotypes does depend on locale (McIntosh & Reinert, 2011), therefore it is unknown whether these two serotypes will become emergent in the Australian population. Besides from this, conjugate vaccines have been designed to target specific pneumococcal serotypes, which is a limitation when determining which serotypes are to be included in the next vaccine. The cost of the vaccine increases as more serotypes are included, hindering the use of these vaccines in developing countries. Therefore alternative vaccines are being investigated.

A cheaper, species-wide protein antigen vaccine is required since the 7vPCV and also 13vPCV are expensive. The expense of the current vaccines particularly impacts on third world countries where IPD is widely prevalent (Kadioglu *et al.*, 2008; Saha *et al.,* 1997; Yaro *et al.,* 2006). Therefore the use of proteomic screening to identify target antigens for a pneumococcal vaccine has been investigated (Moffitt *et al.,* 2011). Common antigens for future vaccines (and combinations thereof) are being studied such as pneumococcal surface protein A (PspA), choline binding proteins (PcpA), pneumococcal surface adhesin A (PsaA), non-toxin pneumolysin derivative (PlyD1), metal-binding lipoproteins (PsaA/PiaA), poly-histidine triad proteins (PhtB, PhtD and PhtE), a murein hydrolase (-LytB), and neuraminidase (NanA) (Hicks *et al.,* 2007; Kadioglu *et al.,* 2008; Sharma *et al.,* 2013; Swiatlo & Ware, 2003). A study on rhesus macaques (*Macaca mulatta*) vaccinated with an AS02-adjuvanted PhtD-dPly vaccine demonstrated protection against *S. pneumoniae*-induced pneumonia (Denoël *et al.,* 2011). Another study demonstrated that a trivalent recombinant protein vaccine (PcpA-PhtD-PlyD1) protected infant mice from lethal pneumonia caused by pneumococcal serotypes 6A and 3 (Verhoeven *et al.,* 2014). It was hypothesised that

this trivalent vaccine reduced pneumococcal binding to the epithelium in the airway (Verhoeven *et al.,* 2014). Utilising antigens that mediate $CD4^+$ T cells has been observed to reduce nasopharyngeal carriage of *S. pneumoniae*, although responses vary between adults and children, demonstrating that future vaccines could use the natural clearance mechanisms in the human body (Sharma *et al.,* 2013).

Vaccines that attack more than one species of bacteria have been developed. In 2012 there was a report on the production of a bivalent vaccine that was shown to protect mice from *S. pneumoniae* colonisation and disease, and *Salmonella typhi* which causes typhoid fever (Lu *et al.,* 2012). This bivalent vaccine consists of an antigen SP1572 (pneumococcal protection protein A) fused to non-haemolytic pneumolysoid (Pdt) that is conjugated to a pneumococcal cell wall polysaccharide to target *S. pneumoniae*. It also targets *S. typhi* using another antigen SP2070 (surface-exposed glucose-6-phosphate isomerase) fused to a Pdt that is conjugated to a Vi polysaccharide, an importance virulence factor of *S. typhi*, to elicit functional antibodies against the Vi polysaccharide (Lu *et al.,* 2012). The two antigens in this vaccine were shown to be conserved in all known sequences of pneumococcal strains, and thus should confer protection across a wide range of pneumococci, while the current PCV vaccines are limited to specific pneumococcal serotypes (Lu *et al.,* 2012).

Additionally, another recently developed vaccine injects killed cells of a non-capsulated *S. pneumoniae* which has been shown to illicit a "bifunctional immunity" by causing plasma antibodies to protect against pneumonia as well as IL-IIA-mediated nasopharyngeal clearance (Malley & Anderson, 2012). Other vaccines include live attenuated vaccines, such as SPY1 pneumococcal strain, which has been shown to protect mice from pneumococcal infections (Wu *et al.,* 2014). The SPY1 pneumococcal strain has significantly reduced virulence properties however provides protection against other pneumococcal strains (Wu *et al.,* 2014). Finally, a Phase I trial of a Triple-Protein vaccine which contains PhtD, dPly and nontypeable *Haemophilus influenzae* protein D (PD) has shown that two doses induced humoral immunity and antigen-specific $CD4^+$ T cell responses (Berglund *et al.,* 2014). Increasing IPD associated with existing vaccines could be resolved with these alternative vaccine development strategies, but until then studies of the changing genetic population structure due to the PCVs are required. It is unknown whether these protein-antigen vaccines are also susceptible to genetic variation in the pneumococcal populations. This could be explored by interrogating the available pneumococcal whole genomes.

2.8 Risk factors associated with *S. pneumoniae* diseases: age, gender and race

A number of risk factors have been examined and associated with pneumococcal diseases, particularly age, gender and race. The sex of a child is considered to have little influence on IPD despite higher observed rates in males. This is thought to be more likely associated with other risk factors (Scott *et al.,* 1996). For example, a ratio of 2.05:1 male vs. female children in Bangladesh was observed for pneumococcal disease, however it was thought that the male dominant society meant that male children were more likely to be taken to hospital and therefore reported, leading to the hypothesis that the sex doesn't have an influence on IPD, but gender does (Saha *et al.,* 1997). There have been a number of epidemiology studies that have observed a slightly higher IPD rate in male patients, however the ratio has not been significant or otherwise researchers have failed to comment on the differences (Moore *et al.,* 2015; Scott *et al.,* 1996;).

Age has been known to be an important factor in IPD rates (Singleton *et al.,* 2007), and therefore important to consider in epidemiology studies. IPD cases are largely bimodal in developed countries, usually peaking in children younger than two years and peaking in adults over 65 years, generally due to poorer immune systems (Roche *et al.,* 2003; Scott *et al.,* 1996). However, serotype distribution is different in children younger than five years compared to people over five years (Elberse *et al.,* 2011b; Scott *et al.,* 1996). For example, in The Netherlands the incidence of IPD caused by serotype 1 decreases as age increases; similarly IPD caused by serogroups 6, 14 and 19 abruptly decreases after ten years of age, and serogroup 18 gradually decreases until 30 years (Elberse *et al.,* 2011b). On the other hand, IPD caused by serogroups 3, 8, 7 and 23 increases as age increases (Scott *et al.,* 1996). It was known in 1996 that when children were vaccinated with the *S. pneumoniae* antigens, the maturation of their antibodies varied widely with serotype (Scott *et al.,* 1996). Alternatively some serotypes are more prevalent in adults compared to children, for example Gratten *et al.* (1995) found that serotype 12F was more frequent in Aboriginal adults than in children, however Saha *et al.* (1997) identified 12F (and 7F) predominant in children in Bangladesh in a 1995-1997 study. Torzillo *et al.* (2007) described an outbreak of 12F in 1993 in Australia with no further reports of 12F outbreaks until 2001. Serotype 12F is not included in the 13vPCV. No known further study of 12F in Australia has been published.

Race is also known to be associated with high incidence rates of IPD, however it is important to keep in mind that there is no good evidence for genetic predisposition to pneumococcal disease related to race. The most overwhelming risk factor is exposure to high rates of pneumococcal colonisation in the population; this is associated with factors such as overcrowding (Jacups & Cheng, 2011; Mehr & Wood, 2012). Indigenous patients are known to acquire IPD at a younger age compared to non-Indigenous patients and there is a higher incidence of IPD in Indigenous citizens than non-Indigenous citizens (Hanna *et al.,* 1997; Mak, 2004). Weatherholtz *et al.* (2010) stated that the Navajo citizens in America are five to ten times more likely to contract IPD than the general US population, and Australian Aborigines are four to ten times more likely to be hospitalised with pneumoniae or influenza than non-Aborigines (Hanna *et al.,* 1997). Gratten *et al.* (1995) observed higher IPD cases in Aborigines in North Queensland, and identified risk factors such as alcohol abuse, which was high in these communities. Most Aborigines with IPD have at least one other risk factor e.g. alcohol abuse, other medical condition, etc. (Hanna *et al.* 1997).

There are also differences in rank order of serotypes that cause disease in different ethnic groups. IPD in Australian Aboriginal citizens was mostly caused by serotype 7F (Gratten *et al.,* 1997; Hanna *et al.,* 1997), which has also been described in Native Alaskan adults (Weatherholtz *et al.,* 2010). Fagan *et al.* (2001) identified a greater variety of serotypes causing IPD in Indigenous than non-Indigenous individuals. According to Hicks *et al.* (2007), there are reports of race-specific trends of serotype replacement (however there are no reports on age-specific trends). It is unknown whether there could be genetic variances in Indigenous populations that pre-determine a higher rate of infectious disease.

2.9 Genetics of *Streptococcus pneumoniae* serotypes

A number of hypotheses have been postulated for the increase in non-vaccine serotypes. A phenomenon called serotype replacement describes the increase in non-vaccine serotypes due to an opening in the niche that vaccine-targeted serotypes once occupied (Weinberger *et al.,* 2011). Serotype replacement has been observed in numerous studies worldwide, and has prompted the development of additional vaccines to cover these emerging serotypes. Further discussion of serotype replacement is provided in Section 2.11. Preliminary identification of serotype replacement can be observed through the 'emergent serotype(s)', which is a non-

vaccine serotype that becomes a carrier in >5% of the population unless already exceeding this before introduction of a vaccine (Cohen *et al.,* 2012).

As well as serotype replacement, the pneumococcus has successfully evaded vaccine pressure by utilising its ability to undergo natural genetic transformation (Straume *et al.,* 2014). Frederick Griffiths described the first evidence of transformation in the 1920's when a non-encapsulated pneumococcus incorporated free DNA into its genome to become capsulated in an *in vivo* (mice) study (Griffith, 1928). Oswald T. Avery also demonstrated this *in vitro* in 1930's-1940's (Avery *et al.,* 1944). This stage of transformation occurs when the bacterium becomes competent, meaning that it has the ability to take up DNA and transform (Johnston *et al.,* 2014; Knutsen *et al.,* 2006). All *S. pneumoniae* can become competent but only briefly in the growth stage.

Competence, for general transformation, is regulated by competence stimulating peptide (CSP) through a signal transduction pathway including histidine kinase comD and cognate response regulator comE (Knutsen *et al.,* 2006). This process of taking naked DNA from the external environment and incorporating it into its genome has enabled the uptake of new capsular genes, hence transforming its expressed capsule into a non-vaccine targeted capsule. Termed "capsule switching", it has been observed in several studies. A number of studies have demonstrated that multiple fragments, some up to 44kb in length, are exchanged between pneumococci (Brueggemann *et al.,* 2007; Golubchik *et al.,* 2012; Wyres *et al.,* 2012). Further discussion of capsule switching is provided in Section 2.10. Interestingly, it is thought that BOX elements play an important role in fine-tuning of the development of spontaneous competence, which may play a role in evolution of the pneumococcal genome (Knutsen *et al.,* 2006). BOX elements are also used for genotyping as explained in *Appendix B8: Published manuscript "Genotyping Streptococcus pneumoniae" to Future Microbiology (April 2015)*.

2.10 Pneumococcal capsule switching

Capsule switching is defined as the horizontal transfer of pneumococcal capsular genes from one pneumococcus to another so that even though the genome is very similar to the original, the expressed capsule polysaccharide is different (Brueggemann *et al.,* 2007). Capsule switching is usually identified when two pneumococci have the same background DNA (e.g. same MLST sequence type) but have different capsular types (Hanage *et al.,* 2005). For example, Moore *et al.* (2008) demonstrated through MLST that ST695 (sequence type) and ST899 were both previously serotype 4 but now both had 19A capsules, indicating capsular switching from vaccine serotype to non-

vaccine serotype. The terminology 'capsule switching' has been considered misleading because it connotes that only the capsule cassette is switched, when in reality genes lying outside of the capsule cassette region can also be included in the switch (Golubchik *et al.,* 2012). For this thesis, capsule switching is recognised as a process where capsule genes are exchanged through horizontal transformation so that the expressed capsule changes.

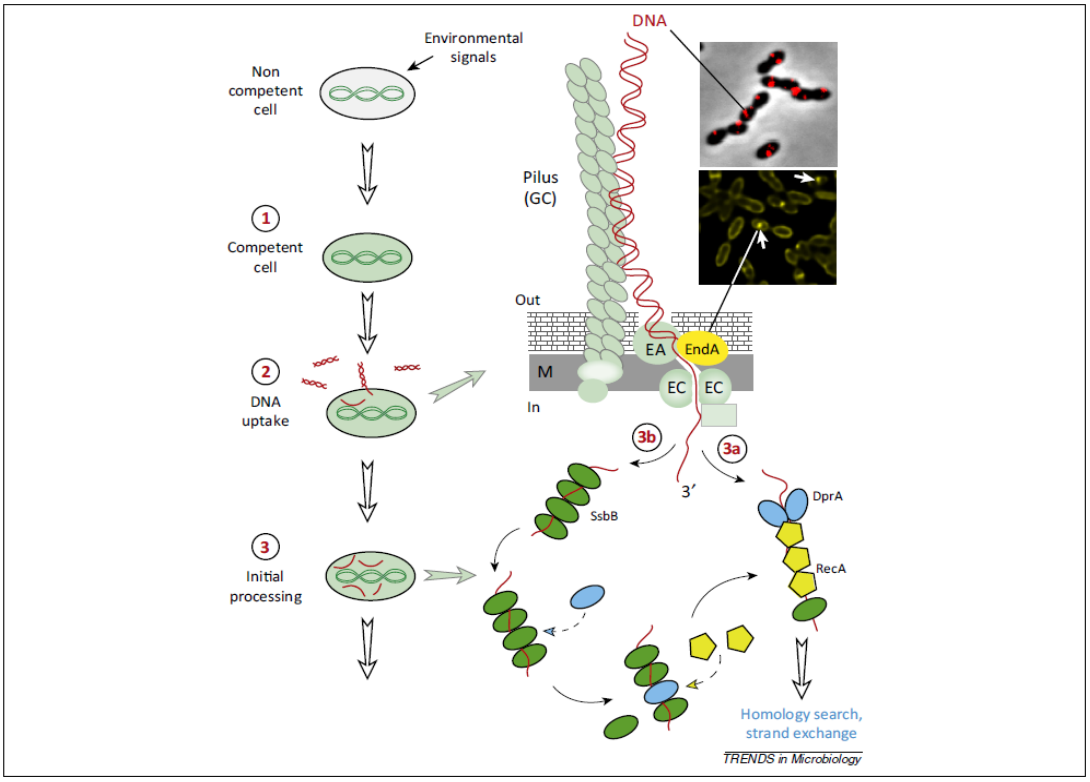
It is important to keep in mind that capsular switching has been observed before the vaccination era. Therefore "natural" capsular switching (not the result of vaccine pressure) should be differentiated from "vaccine-selected" capsule switching (selective pressure of conjugate vaccination) (Beall *et al.,* 2006; Jefferies *et al.,* 2004; Temime *et al.,* 2008). Natural capsule switching has been found to occur at a rate of $1.5x10^{-3}$ per week; however there was not enough evidence to determine the rate of vaccine-selected capsule switching (Temime *et al.,* 2008).

Interestingly, environmental stresses such as antibiotics or vaccines can induce competence in pneumococci, enabling the pathogen to adapt (Johnston *et al.,* 2014). Recent studies have shown that once competent, the pneumococcus can capture foreign dsDNA using a long transformation pilus located mid-cell (Figure 4) (Croucher *et al.,* 2012; Johnston *et al.,* 2014). The dsDNA is then passed to a comEA receptor at the pore complex which transfers the dsDNA to the EndA nuclease to have one strand of DNA degraded (Croucher *et al.,* 2012; Johnston *et al.,* 2014). The resulting ssDNA is cleaved into fragments of approximately 6.6kb in length and crosses the membrane via the ComEC pore (Croucher *et al.,* 2012; Johnston *et al.,* 2014). A nucleoprotein filament is formed when ssDNA is protected by SsdB proteins (RecA and DprA) from further degradation in the cytosol before eventually being incorporated into the genome at regions of similar sequence (Croucher *et al.,* 2012; Johnston *et al.,* 2014).

The total size of alternative coding at the capsular locus is more than 1.8Mbp which is almost the equivalent of an entire *S. pneumoniae* chromosomal complex (Bentley *et al.,* 2006). The capsular genes have low G/C content which may allow new serotypes to form by lateral gene transfer of novel capsule polysaccharide genes (Bentley *et al.,* 2006). The pneumococcus has been known to undergo transformation more frequently than most other bacterial species (Hall *et al.,* 1998).

Figure 4: The process of horizontal transformation in *Streptococcus pneumoniae* via use of a transformation pilus. (Reprinted from *Streptococcus pneumoniae*, le transformiste, Vol. 22 /3, Calum Johnston, Nathalie Campo, Matthieu J. Bergé, Patrice Polard, Jean-Pierre Claverys, *Streptococcus pneumoniae*, le transformiste, Trends in Microbiology: 113-119, Copyright (2014), with permission from Elsevier).

The transformation pilus (GC) captures double-stranded DNA (dsDNA) once the cell becomes competent (1). The dsDNA is passed to the ComEA receptor (EA) before being transferred to the EndA nuclease for degradation. The resulting single-stranded DNA (ssDNA) moves across the membrane via the ComEC pore (EC; 2). SsdB (green ovals) coat the ssDNA to protect it from nucleases, while DprA (blue ovals) also bind to ssDNA to promote nucleation of the recombinase RecA (yellow pentagons) onto the ssDNA. The ssDNA can then be polymerised by RecA, and promote homology search and strand exchange. Green lines, host chromosome; red lines, transforming ssDNA; M, membrane; EA, ComEA; EC, ComEC.



Due to high colonisation of various serotypes in the nasopharynx, this has led to the hypothesis that genetic transfer occurs here (Marsh *et al.,* 2007; Singleton *et al.,* 2007). However, only recently Marks *et al.* (2012) have demonstrated *in vivo* that nasopharyngeal carriage provides extremely efficient (10^{-2}) and high genetic transfer between multiple strains of *S. pneumoniae*. They demonstrated that the SP670 (Pen^r) strain took up genetic DNA from a dead D39-CO8P2 pneumococcal (Erm^r) strain via fratricide (Marks *et al.,* 2012).

Evidence of capsule switching has been recorded worldwide. In particular, MDR pneumococcal isolates are the focus and have been recorded in the PMEN database. The PMEN Spain^{23F}-1 clone has many variants worldwide including serotypes 19F, 14, 19A, 9N and 3 and serogroup 6 (McGee *et al.,* 2001). Similarly Spain^{9V}-3 has many variants including serotypes 14 and 9A and serogroup 19 (McGee *et al.,* 2001). These clones are important to monitor in Queensland since Spain^{23F}-1 has been reported in Australia (as well as Spain^{6B}-2, England^{14}-9 and Taiwan^{19F}-14) (McGee *et al.,* 2001). No other PMEN strains have been reported in Australia, however surveillance is important to determine whether PMEN clones appear and whether capsule switching occurs. Another focus is the MLST online database where isolates of the same MLST profile may have different serotypes. Serotype 19F has the highest strain diversity, indicating that the 19F capsular gene complex is transferred horizontally quite frequently, and this can be seen in the MLST database (Elberse *et al.,* 2011a; Enright & Spratt, 1998). Unfortunately, there is no Australian national database that records detected pneumococcal genotypes and evidence of capsule switching, therefore little has been published in this area.

Most capsule switches have been detected amongst serotypes 19A, 19F, 11A, 15A, 15B and 22F as seen in the MLST database (Aanensen & Spratt, 2005). Serotype 19F was targeted by the 7vPCV, suggesting vaccine pressure has caused this serotype to become readily susceptible to undergo capsule switching. Serotype 19A has been the most successful serotype replacement worldwide, and has been associated with capsule switching e.g. serotype 4 to 19A (Brueggemann *et al.,* 2007). It is unknown why the other serotypes 11A, 15A, 15B and 22F are also commonly observed in switches as they are currently not targeted by any vaccines, however it is not surprising since pneumococci can undergo natural capsule switching. The significance of understanding the ability of *S. pneumoniae* to exchange genetic information will be important when examining capsule switching in the Queensland population. Capsule switching is of major concern particularly if vaccine type serotypes (7vPCV serotypes) transform to non-vaccine capsule types and become "vaccine escape" strains.

The emergence of "vaccine escape" isolates is an alarming possibility of capsule switching. Brueggemann *et al.* (2007) reported that within four years (2003-2007) 37 vaccine escape strains were detected in the USA. It has been shown that multiple and large DNA fragments (sizes ranging from 0.04 – 44kb) were simultaneously transferred in five independent vaccine escape recombination events (Golubchik *et al.,* 2012). Vaccine escape recombinant P1 (ST695 and ST2363) was particularly virulent after switching its capsule from a serotype 4 to a 19A, spreading westward across the USA

after being detected in New York and Connecticut in 2003 (Figure 5) (Golubchik *et al.,* 2012). Pai *et al.* (2005) also described a ST695 serotype 4 isolate with a switched capsule to serotype 19A in USA. In Italy in 2009 and 2010 vaccine escape ST695 was also detected, although it is unsure whether these strains migrated from the USA or whether they independently evolved within the Italian pneumococcal population (Ansaldi *et al.,* 2011). According to the MLST database, a pneumococcus with a MLST profile ST2702, a double locus variant of ST899 serotype 19A (contains two sequence differences in the MLST profile) has been detected in Australia, however it remains as serotype 4 (Aanensen & Spratt, 2005).

Figure 5: Spread of the P1 *S. pneumoniae* vaccine escape strain in the USA during 2003 to 2007. (Reprinted by permission from Macmillan Publishers Ltd: [Nature Genetics] (Golubchik *et al.,* 2012), Copyright (2012))

Bar graphs represent the percentage of serotype 19A isolated within that year; red represents the P1 serotype 19A strain and blue represents other serotype 19A lineages. Years are from 2003 to 2007, left to right across each bar graph. Surveillance in New Mexico began in 2004, hence data from 2003 is missing.

Vaccine escape due to capsular switching is a major concern, especially since the transformation of a serotype can affect the virulence properties. Kelly *et al.* (1994) demonstrated *in vitro* that when a serotype 2 switches to a serotype 3, it had no effect

on virulence; however a serotype 6B strain switching to a serotype 3 became more virulent. Virulence was determined by comparing the serotype 3 derivatives to the parental strains, and measuring the median time to death when female mice were inoculated with the bacteria. Sabharwal *et al.,* (2014) has also demonstrated that pneumococcal virulence properties can change after capsular switching from serotype 6A to 6C. This was demonstrated through infection of mice nasopharynx and detection of viable organisms 48-72 hours later. The significance of identifying capsule switching and vaccine escape isolates in Australia may impact the selection of serotypes for new childhood pneumococcal vaccines.

2.11 Serotype replacement: Prevalence of non-vaccine *Streptococcus pneumoniae* globally

Since the introduction of 7vPCV in the USA in 2000, there has been verified concern that vaccinations will induce serotype replacement. Serotype replacement is "an increase in the incidence of IPD caused by non-vaccine serotypes after the introduction of a vaccine" (Weinberger *et al.,* 2011). It is a gradual and steady process that has been observed all over the world including the USA (Hicks *et al.,* 2007; Reingold *et al.,* 2005; Whitney *et al.,* 2003), Australia (Hanna *et al.,* 2008; Lehmann *et al.,* 2010; Marsh *et al.,* 2007), The Netherlands (Elberse *et al.,* 2011a, Elberse *et al.,* 2012), Portugal (Nunes *et al.,* 2005), Alaska (Singleton *et al.,* 2007), Spain (Guevara *et al.,* 2009), Finland (Hanage *et al.,* 2005), Scotland (McChlery *et al.,* 2005), Canada (Tyrrell *et al.,* 2009) and other countries such as The Gambia, Israel, South Africa, France and Greece (Weinberger *et al.,* 2011). The rate of colonisation of non-vaccine serotypes increases in vaccinated populations, therefore serotype replacement is observed mostly in first world countries where the vaccines are readily available and administered (Mehr & Wood, 2012; Weinberger *et al.,* 2011).

The introduction of pneumococcal conjugate vaccines is thought to have created an ecological niche resulting in serotype replacement (Hicks *et al.,* 2007). In the USA insignificant increases were reported in 2003, three years after the introduction of 7vPCV in 2000 (Reingold *et al.,* 2005) but by 2004 non-vaccine serotypes (namely 3, 15, 19A, 22F, and 33F) caused 88% of IPD cases in American children under five years; however this remained minor compared to the overall decrease in IPD (Hicks *et al.,* 2007, Reingold *et al.,* 2005). By 2007, Singleton *et al.* (2007) reported a 140% increase of IPD caused by non-vaccine serotypes in Native Alaskan children younger than two years old, six years after the introduction of 7vPCV. Predominant serotypes were 3, 6A,

7F, 8, 12F and 19A (Singleton *et al.,* 2007). Globally serotype 19A has been the predominant non-vaccine serotype causing IPD (Hicks *et al.,* 2007). However, Canada had a more significant increase in serotype 5, with a total increase of 236.3% non-7vPCV serotypes causing IPD from 2002 (7vPCV introduction) to 2006 (Tyrrell *et al.,* 2009). In contrast, no overall increase in non-7vPCV serotypes causing IPD has been observed in the Navajo population in the USA apart from specific serotypes (1, 3, 7F and 19A) (Weatherholtz *et al.,* 2010). It is hypothesised that the observed difference in the Navajo population may be because before the 7vPCV trial, only 50% of IPD were caused by serotypes targeted by this vaccine, and the introduction of the vaccine into this population was gradual (Weatherholtz *et al.,* 2010). Serotype replacement would not have been observable since a majority of IPD cases were already caused by serotypes not targeted by the 7vPCV.

There are regional differences in non-vaccine serotype emergence patterns (McIntosh & Reinert, 2011). Serotypes that have low carriage rates (such as serotype 1) have different strains that are found only in certain areas of the globe. For instance, serotype 1 ST306 is found in continental Europe but serotype 1 ST227 is found in England (Brueggemann & Spratt, 2003). Due to low carriage rates of serotype 1, the strains do not spread throughout populations as quickly. This phenomenon has also been observed by Sandgren *et al.* (2004). On the other hand, serotypes that have high carriage rates (such as 19F) have spread across the globe with strains found in many countries, however these serotypes don't frequently cause IPD (Sandgren *et al.,* 2004). Some serotypes are also more highly clonal than others, meaning that most isolates have the same genotype and have not undergone a lot of genetic variation. For example serotype 7F has been found to be highly clonal with the dominant sequence type ST191 found in Australia, Scotland, Sweden and the USA (Jefferies *et al.,* 2004; Sandgren *et al.,* 2004). Similar reports of serotype replacement in Australia are discussed in the following section.

2.12 Serotype replacement: Prevalence of non-vaccine *Streptococcus pneumoniae* in Australia and Queensland

Serotype replacement has been detected in Australia, mirroring observations worldwide. Initially, since the introduction of 7vPCV to Indigenous children in 2001, epidemiology studies focused on this population. From 1996 to 2004, Giele *et al.* (2007) determined that there was no significant increase non-7vPCV serotypes causing IPD in Indigenous children less than five years in Western Australia (WA). This may be

because the 7vPCV had low coverage (54%) in WA compared to other Australian states, and that the number of isolates in their study was too few to identify changes in specific serotypes (Giele *et al.,* 2007). Hanna *et al.* (2006) reported that in North Queensland, serotypes 6A and 19A were the leading non-vaccine serotypes causing IPD in Indigenous children under five from 2002-2004, a period where only Indigenous children were vaccinated with 7vPCV. There was no observation that these serotypes 6A and 19A were serotype replacements as yet, but it was recommended that further surveillance was required (Hanna *et al.,* 2006). By 2005 when the 7vPCV was freely available for all children, IPD cases caused by non-7vPCV serotypes remained unchanged although there were decreases in 7vPCV serotypes (Roche *et al.,* 2007; Roche *et al.,* 2008). A slight increase of serotype 19A in non-Indigenous children under five in 2006 was observed but was considered to be natural fluctuations (Roche *et al.,* 2008).

However, between 2006-2009 the 11 most common serotypes in north Queensland were serotypes 1, 3, 4, 6A, 7F, 9V, 14, 19A, 22F, 23F and 38 of which only 4, 9V, 14 and 23F were targeted by the 7vPCV (Hanna *et al.,* 2010). There was also a gradual increase in non-vaccine serotypes in the Indigenous population in North Queensland, the dominant serotypes being 6A and 33F in Indigenous children, and 22F, 10F and 18A in Indigenous adults (Hanna *et al.,* 2008). Similarly, Giele *et al.* (2009) stated that from 2001 to 2007 there was an increase in 19A (from 2.5% to 20%) in non-Indigenous Australians in North Queensland.

Interestingly, Williams *et al.* (2011) stated that were was no observable increase in serotype 19A in Indigenous children under two years from 2002 to 2007, although this may be because there were lower number of isolates collected from Indigenous children (n=148 compared to n=1441 non-Indigenous children) in this study. Also the incidence of non-7vPCV serotypes remained 10 times higher in Indigenous children than non-Indigenous children, indicating the effects of the 7vPCV would not be as pronounced in the Indigenous population. Overall, significant increases in non-7vPCV serotypes were observed in non-Indigenous children (from 9.7 – 15.7 per 100 000, $P<0.001$), particularly serotype 19A which caused 38% IPD by 2007 (Williams *et al.,* 2011). Similarly, Lehmann *et al.* (2010) observed in WA that there was an increase in serotype 19A from 1997 to 2007 in both Indigenous children (8 to 12.6 per 100 000) and non-Indigenous children (0.6 to 4.9 per 100 000). Overall, non-7vPCV serotypes had increased in WA Indigenous children (40%) and non-Indigenous children (74%) by 2007 (Lehmann *et al.,* 2010).

There are no published studies on the impact of the 13vPCV in Australia (introduced 2011), although Torzillo *et al.* (2007) reported that the 13vPCV serotypes accounted for 80% of IPD isolates in central Australia, indicating it should provide good coverage. Hanna *et al.* (2010) reported that 22F has emerged as a dominant non-vaccine serotype in North Queensland adults. Serotype 22F is not covered by any of the current vaccines. Other serotypes to monitor include 3, 7F and 33F, although serotypes 3 and 7F are targeted by the 13vPCV (Williams *et al.,* 2011). Collins *et al.* (2013) identified that ten years after the 7vPCV introduction, 7vPCV accounted for a small proportion of carried pneumococci in Australian Aboriginal children, and that many non-13vPCV serotypes are in circulation. Reingold *et al.* (2005) emphasised the need to continue surveillance studies to ensue vaccination strategies encompass all of the present and emerging invasive serotypes and thus reduce and maintain low levels of IPD.

2.13 *Streptococcus pneumoniae* epidemiological surveillance

The increasing emergence of drug resistant strains and the phenomenon of serotype replacement and capsular switching have demonstrated the importance of epidemiological surveillance (Hall *et al.,* 1998; Hermans *et al.,* 1995). Molecular typing profiles are used to study transmission routes of infectious diseases, assess sources of infections and assess the impact of human intervention (i.e. vaccinations/antibiotics) (Schouls & Van der Heide, 2012). As there are 98 serotypes of *S. pneumoniae*, highly informative tools for analysis are required. To accurately assess vaccine impact, it is recommended that data are required two years pre-vaccine introduction and five years post-vaccine introduction (Murray *et al.,* 2014). Epidemiology studies have been useful in the past when determining which serotypes to include in the new vaccine i.e. serotype 19A in the 13vPCV after increase of IPD caused by this serotype. Alternatively advice can be given when a vaccine is not required e.g. the 10vPCV was only introduced into Northern Territory, Australia since serotype 5, which was a newly added in this vaccine, was not prevalent elsewhere in Australia. The first beginnings of a pneumococcal surveillance program began in Queensland in 1989 at the Acute Respiratory Infections Unit, Centre for Public Health Sciences, Queensland Health, Brisbane (Gratten *et al.,* 1997). Traditional characterisation methods for epidemiology studies, such as serotyping (Quellung of Neufeld) only provide a broad picture of changes and cannot differentiate genotypes or capsule switching (Neufeld, 1902). Serotyping methods require multiple Factor serum to distinguish, for example, serotype 6C (first described in 2007) from serotype 6A, and more Factor serum are

being developed as more pneumococcal serotypes are identified (Hanna *et al.,* 2010). However this method is very expensive compared to molecular typing methods, and currently has the ability to distinguish up to 91 types. Further discussion of this method is provided in Section 2.16.

The suitability of genotyping tools for epidemiology studies depends on its ability to distinguish different fingerprint patterns for unrelated isolates from identical patterns for related isolates (Hermans *et al.,* 1995). The current 'gold standard' genotyping method for *S. pneumoniae* is Multi-Locus Sequence Typing (MLST). MLST provides high discriminatory powers, good resolution and good reproducibility; however it also has limitations as described in *Appendix B8: Published manuscript "Genotyping Streptococcus pneumoniae" to Future Microbiology (April 2015)*. Another promising genotyping method has emerged, namely Multi-Locus Variable Number Tandem Repeat Analysis (MLVA), also discussed in *Appendix B8: Published manuscript "Genotyping Streptococcus pneumoniae" to Future Microbiology (April 2015)*. Unfortunately Australian states are limited when comparing epidemiology studies due to lack of a national database, and because of this no nationally accepted genotyping method is used. Most Australian epidemiology studies still favour Quellung serotyping and carriage surveillance.

2.14 Capsule typing *Streptococcus pneumoniae*

Characterisation of bacteria below the species level enable the assessment of the impact of human interventions such as vaccines and antibiotics, the relatedness of bacterial isolates, and the sources and transmission routes of infections (Elberse *et al.,* 2011a). Traditional characterisation techniques, such as Quellung serotyping, has been used for decades in pneumococcal epidemiology studies (Neufeld, 1902). Unfortunately, less than 98 published serotypes can be determined when using serotyping, compared to thousands of genotypes identified within these 98 serotypes when using a molecular based method such as MLST. As well as this, not all serotypes can be distinguished, and capsule switching cannot be detected without the assistance of a genotyping method. Capsule typing of *S. pneumoniae* based exclusively on DNA-based methods has several advantages but also has the limitation that capsular expression is not really accessed. For example, the entire capsular operon may be present but due to mutations no capsule is being produced (Salter *et al.,* 2012). A number of molecular based methods using PCR to directly type the capsule cassette sequence have been developed as discussed in the following sections.

2.15 *Streptococcus pneumoniae* capsule cassette

Using PCR methods to determine the capsule type of *S. pneumoniae* relies on the knowledge of the polysaccharide capsule and its associated genes. Almost all 98 serotypes have had the capsule locus (cassette) sequenced, and cassette lengths vary from 10kb to 31kb (Bentley *et al.,* 2006). There has been an estimated 1.8Mbp of alternative capsule coding sequence in the *S. pneumoniae* genome (Bentley *et al.,* 2006). The general structure of the capsule cassette (*cps*) includes four highly conserved translocation and regulation genes at the 5' end (*wzg, wzh, wzd* and *wze*), often followed by a gene *wchA* that encodes for the initial glucose phosphate transferase (Bentley *et al.,* 2006) (Figure 6). A polysaccharide polymerase gene *wzy* and flippase gene *wzx* are often present downstream with a varying set of other genes for acetyl transferase, glycosyl transferases, nucleotide diphosphate, sugar biosynthesis and modifying enzymes (Bentley *et al.,* 2006).

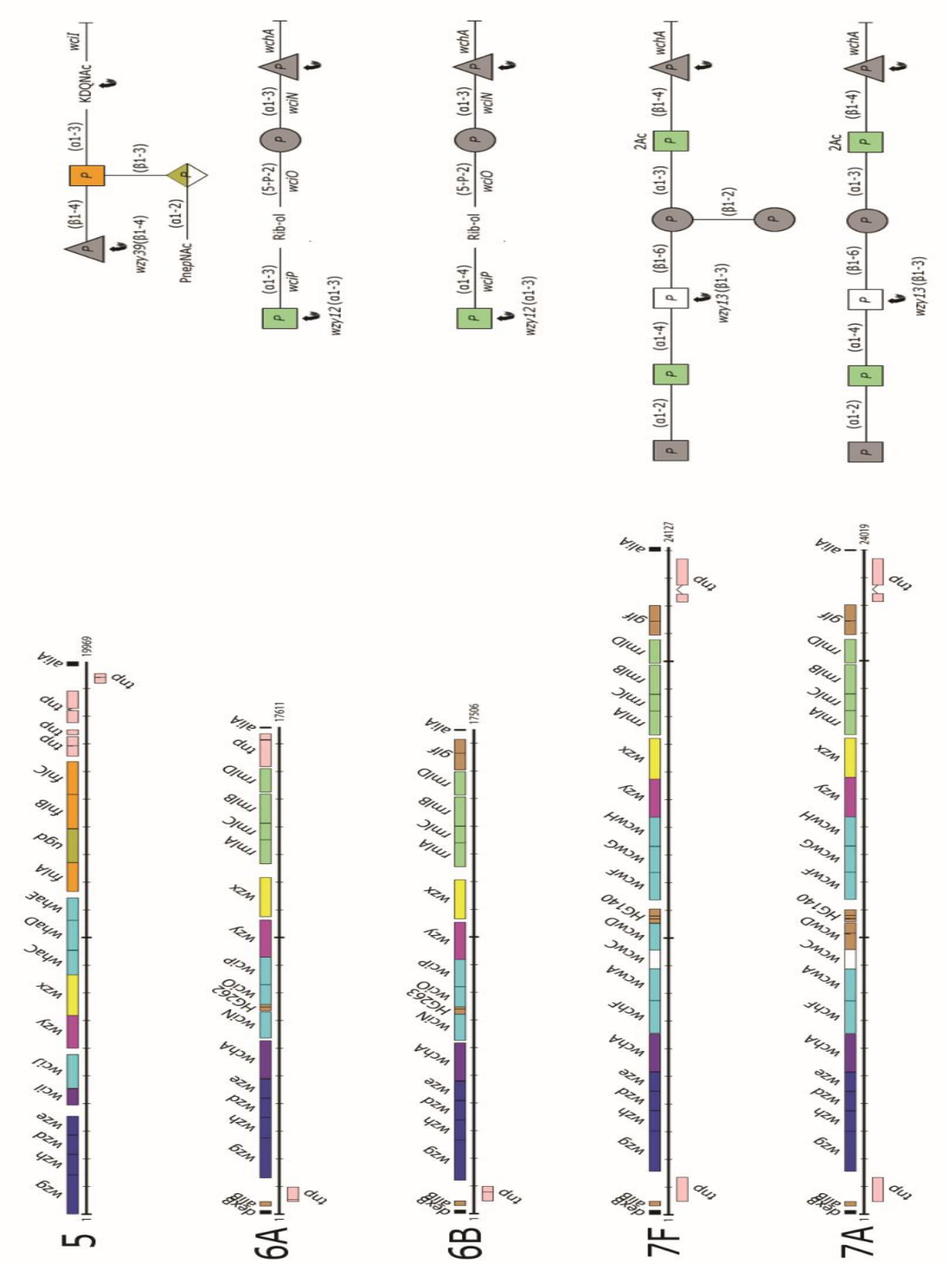
The complete capsule cassette is located between conserved regions *dexB* and *aliA* except for serotypes 3 and 37. In these two serotypes a *wchE* synthase gene is defect; therefore the *tts* gene located elsewhere in the genome is active, replacing the function of *wchE* (Bentley *et al.,* 2006). The region between the *dexB* and *aliA* genes and the rest of the *cps* locus contain mobile genetic elements which may account for some of the capsule switching (Bentley *et al.,* 2006). The entire capsule cassette has a lower G/C content than the rest of the genome (Bentley *et al.,* 2006; Mavroidi *et al.,* 2007).

Published data on already sequenced *S. pneumoniae* capsule loci show that differences between serotypes can be major or minor. For example, the only difference between serotype 6A and serotype 6B are SNP differences in the *wciP* gene, whilst serotype 6C has a unique *wciN* gene and serotype 6D has a $wciP_{\beta}$ gene (Bratcher *et al.,* 2011). The $wciP_{\beta}$ is an allele associated to serotype 6B and 6D, while the $wciP_{\propto}$ allele is associated to serotype 6A and 6C; this can be used to distinguish these two sub-groups of serogroup 6 (Bratcher *et al.,* 2011).

The ability to switch the capsule genes is a concern because the current childhood vaccines (7vPCV, 10vPCV and 13vPCV) act to target the polysaccharide capsules of those serotypes included in the vaccine (i.e. for 13vPCV it targets serotype 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F and 23F).

Figure 6: Example of *S. pneumoniae* capsule cassettes and polysaccharide structure for serotypes 5, 6A, 6B, 7F and 7A (adapted from Bentley *et al.,* 2006)**.**

Gene designations are indicated above each coloured box, and genes are presented on forward and reverse strands. Polysaccharide structures (right side) are coloured by grey, housekeeping sugars; orange, *N*-acetylfucosamine; green, rhamnose; white, *N*-acetylgalactosamine.



2.16 Molecular capsule typing

Since capsule switching has been identified as an important phenomenon to investigate, particularly due to vaccine escape serotypes, several techniques for sequencing the capsule locus have been published. Traditional serotyping methods such as Quellung reaction do not distinguish between all serotypes e.g. between 20A and 20B, or otherwise require multiple Factor serum. PCR and sequencing methods to determine serotypes are more specific and less expensive (Pai *et al.,* 2006). Serotyping only detects the present capsule – it cannot determine what capsule type the bacterium previously had. Serotyping combined with a genotyping method can give preliminary identification that a capsule switch has occurred if given the history of the clone or genotype; however a more accurate detection would require detecting the specific switch within the capsule cassette sequence. Bentley *et al.* (2006) determined the capsule sequences of 90 serotypes using the Expand Long Template PCR System from Roche which sequences up to 20kb. A similar technique has also been used by Brueggemann *et al.* (2007) and Ansaldi *et al.* (2011). However less expensive and faster PCR techniques have been developed to determine capsule types. One of the earliest multiplex PCR methods was developed by Lawrence *et al.* (2003) for determining serotypes 1, 3, 6, 14, 19F and 23F. Primers were combined in 3 multiplex reactions and fragment sizes of PCR products were separated by capillary electrophoresis.

Pai *et al.* (2006) developed a sequential multiplex PCR based on amplifying short regions within the capsule cassette for serotypes common in the Burkina Faso and Togo regions. A seven multiplex PCR with 29 primer pairs could identify 17 serotypes. This sequencing multiplex has been widely applied and optimised to regional pneumococcal populations around the world. Ahn *et al.* (2012) modified this multiplex PCR to a 35 primer pair. Both these methods still require gel electrophoresis to detect the amplification products.

Elberse *et al.* (2011c) also used a different long-PCR method to sequence the capsule locus of serogroup 6 and serotype 19A and 19F, and also created an allele-specific PCR (e.g. sequence only *wzy* and *rmlC* for serogroup 6) to characterise a batch of isolates. The genes *wzg, rmlA* and *rmlB* had the highest degree of sequence variation (Elberse *et al.,* 2011c). Beside from differences between serotypes (e.g. there is a SNP difference in *wciP* gene to distinguish serotype 6A from 6B), capsular subtypes were also found within each serotype (Elberse *et al.,* 2011c). There are six subtypes of serotype 6A, three subtypes of serotype 19A and four subtypes of serotype 19F

(Elberse *et al.,* 2011c). Elberse *et al.* (2011c) theorised that 6C was a descendent from 6A subtype II (6A-II), as well as suggesting that serotype 19A-II may have enhanced polysaccharide production due to a difference in the *wzg* locus compared to 19A-I, particularly since 19A-II has been more successful in vaccinated children. By examining these subtypes, a further understanding of capsule switching and the evolution process can be determined. Elberse *et al.* (2011c) has demonstrated the proposed idea to sequence the complete *cps* cassette and then select polymorphic genes within the cassette to identify potential subtypes. However, this published method lacks details on the allele-specific PCR procedure to identify capsular subtypes. Information for sequencing the complete capsular gene is also lacking compared to other published long-template methods.

Wyres *et al.* (2013) has also demonstrated evolution of the *cps* locus among pneumococcal clonal complexes (CC) of a particular serotype (contains genetically related pneumococci based on their genotypes), for example a serotype 7F in CC191 can have a substitution in the *wzg* locus (7F/5^{b}) or instead in the *wchF* locus (CDC1087-00). However since they have used whole genome sequencing, this is not a practical and cost-effective method for this project particularly since we are only interested in a small section of the whole *S. pneumoniae* genome.

Bratcher *et al.* (2011) and Mavroidi *et al.* (2004) have also examined serogroup 6 using an allele-specific PCR. Bratcher *et al.* (2011) identified differences in the *cps* cassette, which has allowed them to examine the possible evolution of serotype 6C from 6A. They also used allele-specific sequencing of *wciP, wzy* and *wzx* for serogroup 6 (Bratcher *et al.,* 2011). However since they have only examined serogroup 6, these allele-specific PCR primers may not be suitable for other serotypes, particularly if these alleles aren't present in other serotypes. This thesis will examine all the published *csp* loci and determine gene-specific PCR and allele-specific PCR for these loci.

2.17 Summary and Implications

S. pneumoniae is a common but deadly bacterium that causes high mortality and morbidity worldwide. Despite decreases in IPD after the introduction of the 7vPCV, epidemiology studies have been deemed extremely important particularly due to the shifting genotype population to non-vaccine serotypes. Two phenomenon, serotype replacement and capsule switching, have been the main reasons for the changing population structure. *S. pneumoniae* populations differ within age groups, geographic locality and ethnicity; therefore it is vital that studies are continued in each area.

Although Australia is considered as a first-world country, the Australian Aborigines retain the highest rates of pneumococcal diseases in the world. A 13vPCV was introduced for Australian children under two in July 2011, targeting an extra six pneumococcal serotypes that was not included in the 7vPCV. Studies are emerging examining the effect of the 13vPCV in different countries, however currently no studies have been published investigating whether the genetic population has changed in Australia.

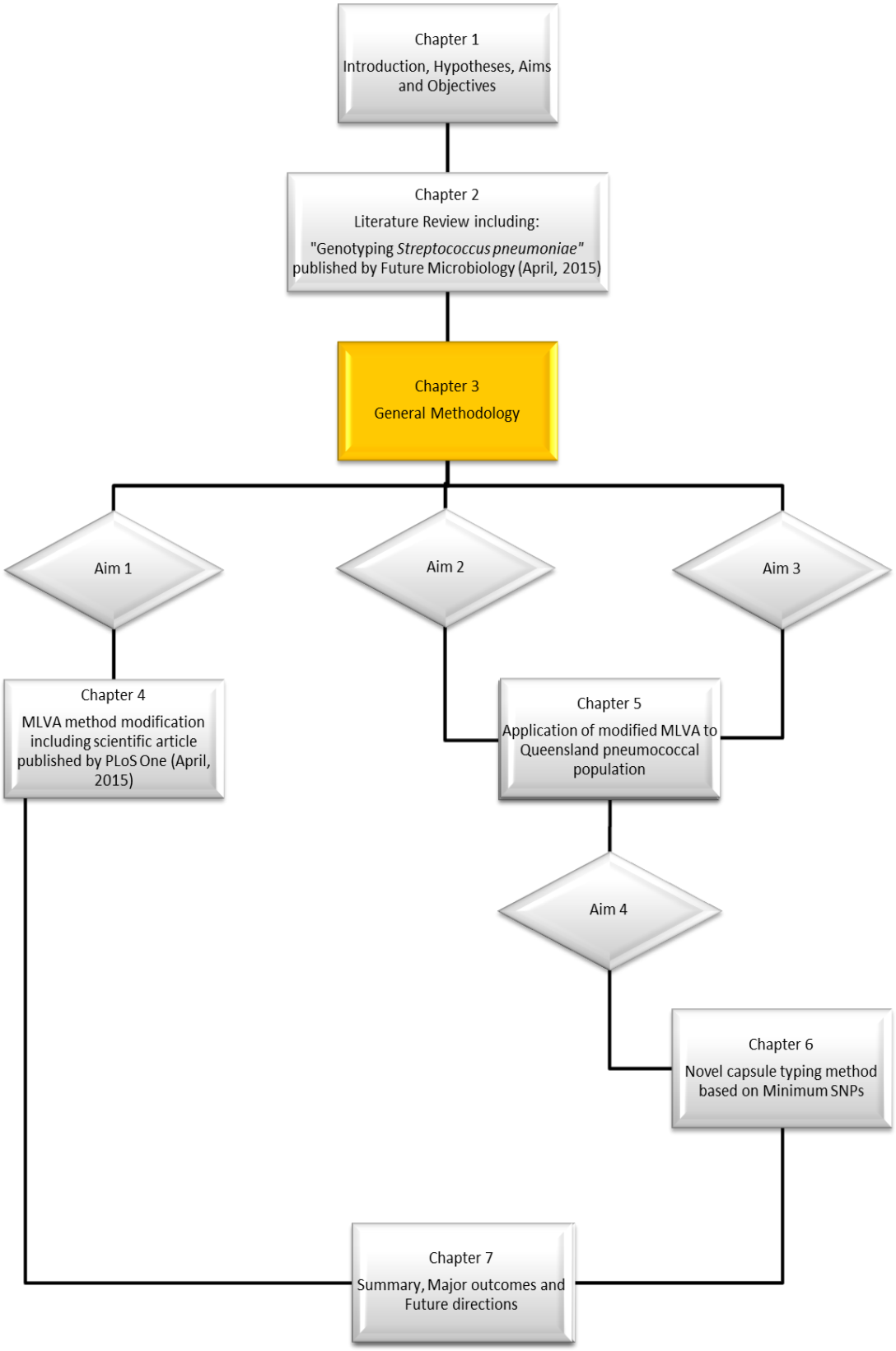
Genotyping *S. pneumoniae* is important for pneumococcal epidemiology studies. By adopting genotyping methods in conjunction with serotyping methods, further genetic information of invasive pneumococcal strains can be obtained. The most favourable genotyping methods ideally have high throughput, high discrimination, good typeability and low costs. There are a number of different genotyping methods available, however MLVA is emerging as a highly discriminatory, quick and inexpensive method compared to the 'gold standard' MLST for genotyping *S. pneumoniae.* There have been at least four different MLVA methods published for genotyping *S. pneumoniae,* each method containing different targets in the pneumococci genome. Unfortunately, no universal MLVA method currently exists and limitations remain such as non-amplified genotyping profiles, therefore further study is required so that comparison of pneumococcal MLVA types between countries can be easily achieved.

The future direction of genotyping methods may continue to see the combination of many genes, including BOX loci and housekeeping genes, with mass spectrometry (MS) or Next Generation Sequencing (NGS). The benefits of these new technologies (MS and NGS) may decrease cost and time. The choice of VNTR or BOX loci raises the question of whether they are stable *in vitro* and *in vivo,* particularly since it is known that *S. pneumoniae* has high mutation rates within the genome. Published literature has demonstrated that these genes are stable in various conditions, although it is unknown whether all VNTR loci remain stable depending on how many times and in what manner the clinical isolates have been sub-cultured before they reach a research laboratory.

As well as methods for genotyping *S. pneumoniae*, investigations examining capsule typing have been performed. The *S. pneumoniae* polysaccharide capsule plays an important role in virulence, and is the target of a number of pneumococcal vaccines that are implemented in national immunisation schemes worldwide. Multiple and large fragments of DNA have been reported to undergo recombination, resulting in vaccine escape strains in the USA and Italy. This is a major concern since vaccine escape strains have the ability to become immune to the pneumococcal vaccine. Traditional

serotyping methods, mainly the Quellung reaction, allow the detection of pneumococcal serotypes. However serotyping is expensive, laborious and cannot distinguish all serotypes, therefore a number of molecular based capsular typing methods have been developed. Capsular sub-types within each serotype have also been detected and associated with the emergence and persistence of certain pneumococcal strains after the introduction of the vaccines. Currently, no molecular capsular typing method has been applied to all 98 known *S. pneumoniae* capsule types.

CHAPTER 3: GENERAL METHODOLOGY


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

3.1 Bacterial isolates

In Australia, all cases of IPD caused by invasive *S. pneumoniae* are reported by general practitioners, public and private hospitals to the State Communicable Disease Department in the Department of Health, where data and isolates are collated. The national data is updated annually. All Queensland invasive pneumococcal isolates are required to be sent to the Public Health Microbiology Laboratory at Queensland Health Forensic and Scientific Services. Invasive *S. pneumoniae* is identified as the isolation of the bacteria from normally sterile body specimens such as blood, cerebral spinal fluid, pleural fluid and tissues. There can be a difference between the total number of IPD cases reported and the total number of pneumococcal serotypes reported for each interval and year.

The state of Queensland, Australia, has a population of 4.7 million people spread across an area of ~1.7 million km^2 (Queensland Government, 2015). The National Notifiable Diseases Surveillance System recorded a 7.6 per 100 000 notification rate of confirmed cases of IPD in Queensland in 2012 (NNDSS, 2015).

3.2 Phenotypic methods

Invasive *S. pneumoniae* isolates taken from normally sterile body sites (blood, tissues, cerebrospinal fluid) were serotyped by the Pneumococcal Reference Laboratory, QHFSS, using Quellung reaction (Statens Serum Institut, 2013). Serotyping (Quellung) is mandatory, however further genotyping has only been performed for research purposes. Serotyping was performed by staff at QHFSS.

3.3 Genotypic methods

3.3.1 Preparation of genomic DNA

A total of 394 *S. pneumoniae* isolates were identified as isolated from children under 15 years old (Table 3). Two *S. pneumoniae* ATCC strains, ATCC49619 (serotype 19F) and ATCC6306 (serotype 6A) were also obtained from Cryosite, Australia. These *S. pneumoniae* isolates were cultured (24hrs, 37°C, 5% CO_2) on Horse Blood Agar (HBA, Oxoid Australia) with the 16-streak technique using a 1mm white loop. A single pneumococcal colony was re-cultured on fresh HBA to ensure isolates were genetically identical. A loopful of colonies was boiled in 400µL TE buffer (~pH8.0) for eight minutes to prepare a thermolysate containing template DNA. A boiler or a

Ratek Dry Block Heater (model no. DBH30D, Ratek) was used to boil the samples at 100°C. The thermolysates were stored at -20°C until further use. Thermolysate DNA was used for genotyping work, including MLST, MLVA and the capsule typing method.

Table 3: *S. pneumoniae* isolates, used in this study, were taken from Queensland children 15 years or younger from 2007 to 2012.

Isolates were collected from sterile body sites. No patient information was available (e.g. age, gender, geographic location, vaccination history, etc.)

Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype
1	2011	19F	81	2011	19A	219	2008	33F	319	2007	19A
2	2011	19F	82	2011	19A	220	2008	34	320	2007	19A
3	2011	22F	84	2011	19A	221	2008	35B	321	2007	19A
4	2011	23A	85	2011	19A	222	2008	38	322	2007	19A
5	2011	23B	86	2011	19A	223	2008	4	323	2007	19A
6	2011	23B	87	2011	19A	224	2008	6A	324	2007	19A
7	2011	23B	88	2011	19A	225	2008	6A	325	2007	19A
8	2011	3	89	2011	19A	226	2008	6A	326	2007	19A
9	2011	3	90	2010	1	227	2008	6B	327	2007	19A
10	2011	3	91	2010	1	228	2008	7F	328	2012	1
11	2011	3	92	2010	1	229	2008	7F	329	2012	1
12	2011	3	93	2010	11A	230	2008	7F	330	2012	1
13	2011	3	94	2010	15B	231	2008	7F	331	2012	1
14	2011	3	95	2010	15C	232	2008	8	332	2012	1
15	2011	33F	96	2010	18A	233	2008	8	333	2012	1
16	2011	35B	98	2010	18C	234	2008	9N	334	2012	1
17	2011	38	99	2010	19F	255	2008	22F	335	2012	11A
18	2011	6C	100	2010	19F	256	2008	18C	336	2012	15B
19	2011	6C	101	2010	19F	257	2008	19A	337	2012	19A
20	2011	6C	102	2010	23A	258	2008	19A	338	2012	19A
21	2011	7F	103	2010	23A	259	2008	19A	339	2012	9A
22	2011	7F	104	2010	3	260	2008	1	340	2012	19A
23	2011	7F	105	2010	33F	261	2008	12F	341	2012	19A
24	2011	7F	106	2010	35B	262	2008	12F	342	2012	19A
25	2011	7F	107	2010	6C	263	2008	18A	343	2012	19F

Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype
26	2011	7F	108	2010	6C	264	2008	18C	344	2012	23B
27	2011	7F	109	2010	6C	265	2007	1	345	2012	33F
28	2011	7F	110	2010	6C	266	2007	1	346	2012	6C
29	2011	7F	111	2010	6C	267	2007	1	347	2012	6C
30	2011	7F	112	2010	6C	268	2007	1	348	2012	7F
31	2011	7F	113	2010	7F	269	2007	1	349	2012	7F
32	2011	7F	114	2010	7F	270	2007	1	350	2012	7F
33	2011	7F	115	2010	7F	271	2007	1	351	2012	7F
34	2011	9V	142	2009	19F	272	2007	3	352	2012	35B
35	2011	1	143	2009	22F	273	2007	3	353	2012	7F
36	2011	1	144	2009	23B	274	2007	3	354	2012	7F
37	2011	1	145	2009	23F	275	2007	3	355	2012	1
38	2011	1	146	2009	23F	276	2007	4	356	2012	1
39	2011	1	147	2009	3	277	2007	14	357	2012	3
40	2011	1	148	2009	3	278	2007	14	358	2012	4
41	2011	1	149	2009	3	279	2007	38	359	2012	15C
42	2011	1	150	2009	33F	280	2007	38	360	2012	15C
43	2011	1	151	2009	35B	281	2007	10A	361	2012	19A
44	2011	1	152	2009	35F	282	2007	10A	362	2012	19A
45	2011	1	153	2009	38	283	2007	10A	363	2012	23B
46	2011	1	154	2009	38	284	2007	15B	364	2012	23B
47	2011	11A	155	2009	38	285	2007	15C	365	2012	33F
48	2011	12F	156	2009	38	286	2007	15C	366	2012	33F
49	2011	12F	157	2009	4	287	2007	15C	367	2012	33F
50	2011	15B	158	2009	6A	288	2007	15C	368	2012	35F
51	2011	15B	159	2009	6C	289	2007	15C	369	2012	7F
52	2011	15C	160	2009	7F	290	2007	18A	370	2012	1
53	2011	15C	161	2009	7F	291	2007	18B	371	2012	7F
54	2011	15C	162	2009	8	292	2007	18C	372	2012	7F
55	2011	18A	163	2009	8	293	2007	19F	373	2012	9N
56	2011	18A	164	2009	9N	294	2007	19F	374	2012	34
57	2011	19A	165	2009	9V	295	2007	19F	375	2012	16F
58	2011	19	166	2009	19F	296	2007	19F	376	2012	22F
59	2011	19A	167	2009	1	297	2007	22F	377	2012	19A

Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype	Isolate ID	Year of isolation	Serotype
60	2011	19A	168	2009	1	298	2007	22F	378	2012	4
61	2011	19A	169	2009	10F	299	2007	23A	379	2012	35F
62	2011	19A	170	2009	10F	300	2007	23F	380	2012	19A
63	2011	19A	171	2009	14	301	2007	23F	381	2012	7F
64	2011	19A	172	2009	15B	302	2007	33F	382	2012	14
65	2011	19A	173	2009	18C	303	2007	33F	383	2012	3
66	2011	19A	174	2009	18C	304	2007	6B	384	2012	35F
67	2011	19A	205	2008	19F	305	2007	7F	385	2012	33B
68	2011	19A	206	2008	19F	306	2007	7F	386	2012	6A
69	2011	19A	207	2008	19F	307	2007	9V	387	2012	19A
70	2011	19A	208	2008	22A	308	2007	19A	388	2012	7F
71	2011	19A	209	2008	22F	309	2007	19A	389	2012	6C
72	2011	19A	210	2008	23A	310	2007	19A	390	2012	19A
73	2011	19A	211	2008	23F	311	2007	19A	391	2012	19A
74	2011	19A	212	2008	23F	312	2007	19A	392	2012	7F
75	2011	19A	213	2008	23F	313	2007	19A	393	2013	23B
76	2011	19A	214	2008	24F	314	2007	19A	394	2013	1
77	2011	19A	215	2008	3	315	2007	19A			
78	2011	19A	216	2008	3	316	2007	19A			
79	2011	19A	217	2008	33F	317	2007	19A			
80	2011	19A	218	2008	33F	318	2007	19A			

Pneumococcal colonies from the second HBA plate were also prepared for long-term storage. Colonies were grown for a further 24hrs in Brain Heart Infusion (BHI, Becton Dickenson, Australia) broth at 37^{o}C, 5% CO_2 and were spun down in a centrifuge at 8000rcf for 30min. The broth was carefully removed by pipette and 200µL saline wash was added to the cells. The colonies were mixed using a vortex and centrifuged a second time and the saline wash carefully removed. Cell pellets were left at the bottom of the Nunc tubes (Sigma-Aldrich, Australia) and placed in -80^{o}C freezer until further use.

Extraction of purified *S. pneumoniae* DNA was performed following the QIAamp DNA Mini Kit (50) (Qiagen, Australia). Concentration of extracted DNA was measured on NanoDrop ND-1000 (Thermo Scientific) with genomic yields of 8ng/µL to 50ng/µL. Extraction of purified DNA for two *S. pneumoniae* ATCC strains, *Streptococcus pyogenes*

and *Streptococcus agalactiae* was also performed using the QIAamp DNA Mini Kit. Purified DNA was stored in -20°C freezer until further use. QIAGEN extracted DNA was prepared for samples tested in Chapter 6 regarding the capsule typing method, in case the high resolution melt (HRM) with real-time PCR was not effective when using the thermolysate DNA. HRM is much more sensitive than conventional PCR and the thermolysate DNA may not be of high enough purity and contain PCR inhibitors.

3.3.2 Gel electrophoresis

Gel electrophoresis was used for the visualisation of the amplified PCR products. Agarose Low EEO powder (Roche, Spain) was added to 100mL of TAE or TBE buffer and heated in the microwave for 2min (e.g. 2g agarose for 2% gel). A staining dye, such as ethidium bromide or SYBR® Safe DNA Gel Stain (Invitrogen™, Australia) was added to the heated gel. The gel was poured into a mould with a comb and left to set. Once set, the comb was removed and gel was transferred to an electrophoresis bath and covered with the appropriate buffer (TAE or TBE depending on what the gel was made from). A loading dye was added to PCR samples before pipetting into the gel wells. An appropriate DNA ladder was also used per gel, such as DNA Molecular Weight Marker VIII or XIV (Roche, Germany). PCR products were run usually at 400mA and 80V. Time would vary depending on the gel and the percentage of agarose in the gel (higher percentage agarose gel means longer run time). Gels were visualised under UV light using the GelDoc XR system (Bio-Rad, Australia) or the G-BOX Syngene system (SynGene, England).

3.3.3 MLST genotyping

Isolates of *S. pneumoniae* selected for MLST genotyping included 13vPCV serotypes (except serotype 19A) and non-13vPCV serotypes to minimise labour and costs, but also to allow examination of changes in these vaccine targeted or non-targeted serotypes. MLST was applied to selected isolates for comparison purposes as previously described using the Corbett Cas1200 Robotics (Qiagen, Australia) for automation (Enright & Spratt, 1998; Jefferies *et al.,* 2003). Housekeeping genes for MLST were blasted in NCBI database against five known *S. pneumoniae* strains (R6, Hungary19A-6, CGSP14, G54 and TIGR4) (*Appendix A1: Position of MLST housekeeping primers in five* S. pneumoniae *genomes).*

Seven housekeeping genes (*aroE*, *gdh, gki, recP, spi, xpt* and *ddl*) were amplified using PCR (Table 4). PCR reactions (total volume 28μL) consisted of 10X PCR Reaction

Buffer (Roche, USA), 25mM $MgCl_2$ (Invitrogen, Australia), 20mM PCR Grade Nucleotide Mix (Roche, USA), 5.6ρmol of forward and reverse primers (Geneworks, Australia), Taq DNA Polymerase (Roche, USA) and reverse-osmosis water (PALL, Australia). Thermocycler (Eppendorf, Australia) parameters were as follows: activation of AmpliTaq Gold at 95°C for 9min, 35 cycles of 94°C for 30sec, 55°C for 60sec and 72°C for 30sec, followed by 72°C for 10min and hold at 4°C. PCR products were run on 1.5% agarose gel to check for amplification.

Table 4: MLST housekeeping gene primers for sequencing *S. pneumoniae* (Enright & Spratt, 1998).

Primer	Forward	Reverse
aroE	5'-GCCTTTGAGGCGACAGC	5'-TGCAGTTCA(G/A)AAACAT(A/T)TCTAA
gdh	5'-ATGGACAAACCAGC(G/A/T/C)AG(C/T)TT	5'-GCT TGA GGT CCC AT(G/A) CT(G/A/T/C) CC
gki	5'-GGC ATT GGA ATG GGA TCA CC	5'-TCT CCC GCA GCT GAC AC
recP	5'-GCC AAC TCA GGT CAT CCA GG	5'- TGC AAC CGT AGC ATT GTA AC
spi	5'-TTA TTC CTC CTG ATT CTG TC	5'-GTG ATT GGC CAG AAG CGG AA
xpt	5'-TTA TTA GAA GAG CGC ATC CT	5'-AGA TCT GCC TCC TTA AAT AC
ddl	5'-TGC (C/T)CA AGT TCC TTA TGT GG	5'-CAC TGG GT(G/A) AAA CC(A/T) GGC AT

The amplified DNA fragments were washed with exoSAP solution (X10 Buffer, 5U/μL phosphatase, 20U/μL exonuclease and reverse osmosis water (PALL, Australia)) to eliminate un-incorporated primers and dNTPs. ExoSAP was then inactivated by heating at 37°C for 45min, then 80°C for 15min, and held at 4°C. A quantitative gel check was performed to determine the suitable amount of PCR product required for sequencing.

Appropriate dilutions of the PCR+exoSAP (inactivated) were made according to the quantitative gel. BigDye fluorescent label was incorporated to the PCR+exoSAP products during a second PCR run of 95°C for 30sec, 25 cycles of 95°C for 10sec, 50°C for 5sec and 60°C for 4min, before hold at 4°C. Fluorescently labelled PCR products were transferred to 96-well plate for purification.

A plate washing protocol was performed as follows: 52μL of a 95% ethanol (7mL fresh) and 3M NaOAc (ph4.6; 280μL) solution was added to each well, vortexed and incubated at room temperature for 45min; the sealed plate was centrifuged at 4500rpm for 60min; the seal was removed, the plate inverted and placed upside down on paper towels in a centrifuge before spinning at 1100rpm for 1min; the plate was removed and 150μL of 70% ethanol (fresh) was added, vortexed and centrifuged upright and sealed at 4500rpm for 10min; the plate was inverted on paper towels in the centrifuge and spun again at 1100rpm for 1 min; the plate was removed and allowed to dry in a fume cupboard before sealing and storing at -20°C.

Before sequencing on the AB3130 DNA sequencer (Applied Biosystems, Australia), the plate was resuspended with 10μL formamide. A denaturing step was performed by heating the plate to 95°C for 5min before quickly centrifuging to get rid of any air bubbles. The plate was placed in AB3130 for sequencing.

3.3.4 MLVA primer design and multiplexing

MLVA4 primers were re-designed to improve annealing for BOX-10, BOX-12, BOX-13 and Spneu19 using Oligo v6.57 software (Wojciech Rychlik). Primer dimer check was performed using PriDimerCheck (Qu *et al.,* 2009), which is available from http://biocompute.bmi.ac.cn/MPprimer/primer_dimer.html. From this it was determined that certain primers would not be combined in the same multiplex reaction due to primer-dimer formation. All MLVA primers and newly re-designed primers were blasted against five known *S. pneumoniae* strains (strain R6, Hungary 19A-6, G54, CGSP14 and TIGR4) in the NCBI database (*Appendix A2: BOX primer position in five* S. pneumoniae *genomes)*. These five strains were chosen since R6 and TIGR4 have been used in other published papers, and the other strains are vaccine targeted serotypes, specifically serotype 19A (Hungary19A-6), serotype 19F (G54) and serotype 14 (CGSP14).

Multiplex Manager v1.2 software (Holleley & Geerts, 2009) was used to determine the optimal combination of MLVA4 primers and the minimum number of multiplexes.

The Tandem Repeat Finder algorithm was used for detecting tandem repeats in the NCBI *S. pneumoniae* sequences, particularly when designing the new primers, available from http://tandem.bu.edu/trf/trf.html (Benson, 1999). It was also used to verify the number of tandem repeats expected to be seen *in silico* when using the published primers. Tandem repeats were found under the following parameters: alignment parameter (match, mismatch, indels) = 2, 7, 7; minimum alignment score = 50; maximum period size = 50.

3.3.5 MLVA genotyping methods

The MLVA1 calibration set was sent from researcher Karin Elberse (National Institute for Public Health and the Environment, The Netherlands), to verify the correct lengths of the alleles on our DNA sequencer.

Non-fluorescent MLVA1 primers (Geneworks, Australia) were diluted in 500μL of TE buffer and placed on an orbital mixer for 1 hour. Fluorescently labelled MLVA1

primers (Table 5) (Applied Biosystems, Australia) at 5' end with F (FAM), N (NED), V (VIC) or P (PET) were also diluted in 500µL of TE buffer. Primer concentration was measured by NanoDrop 1000 (Thermo Scientific, Australia) at 260nm, ssDNA-33.0. A working dilution at 10ρmol/µL was used.

A mastermix was made consisting of MLVA1 primers, HotStarTaq Master Mix (Qiagen, Australia) and reverse osmosis water (PALL, Australia). The Corbett Cas1200 Robotic system (Qiagen, Germany) was used to combine 18µL of mastermix to 2µL of *S. pneumoniae* thermolysate DNA into 96-well plates. A protective film was used to seal the wells. The PCR protocol was optimised to 15min at 95°C, 30 cycles of 95°C for 30sec, 58°C for 60sec and 72°C for 60sec, followed by extension of 72°C for 10min and held at 4°C. Gel electrophoresis was performed to verify amplification of PCR products.

Table 5: MLVA1 primer sequences for genotyping *S. pneumoniae* (Elberse *et al.,* 2011a).

	Primer	Forward	Reverse
Multiplex 1	BOX-01	CCAGAGACATTGATGAAGAGA – F	CGCTTTGATGAACTTGAGTT
	BOX-02	TTGCTTGGTACAGAAAACAAA – N	CCCCATAAAACCCTCCTTATA
	BOX-03	TCCAACACGACCTTTATCC – V	TTCAGTAAACCCAGCTCGTA
	BOX-04	a)TGGGTAAAAGTAGACAGGACT b)AGGGGATTTACCCACTACAAA	CACTTCTACACTAGTTTGTAAAGCA – P
Multiplex 2	BOX-06	TTATGATTTTTGTCATTTTGCA	TCACTTGAGACAATCAGCC – N
	BOX-11	TCCAATAATGACAGGTTTTCCTC – V	TTCCAATCTACGCCTTTGAAG
	BOX-12	TTGCCCTTTTCATCTTCGA – P	CAGCAACCATTGAAACGC
	BOX-13	TCGCCTTTGCTAATCAAACA - F	CTGATTATATCGCTCACAAACCC

PCR products were diluted (1:150) semi-automatically using an eight-channel pipette with reverse-osmosis water (PALL, Australia). The use of formamide instead of water was attempted as other published papers had used this for the dilution of amplified products, however poor results were obtained and using water instead produced better results. Furthermore, poor results when using formamide has been noticed by other researchers and additional limitations are that formamide degrades DNA when stored for some time, has a limited shelf life and when shelf-life is reached there is a dramatic drop in efficacy of the formamide to denature the DNA (L. Schouls, personal communication, 2012).

The GeneScan™ 1200LIZ® dye Size Standard internal ladder (Life Technologies, Australia) was diluted (1:30) with reverse-osmosis water (PALL, Australia), and 10µL of diluted ladder was combined with 2µL of diluted PCR product. A heat denaturation step of 95°C for 5min, followed by a hold step at 4°C was performed on a thermocycler (Eppendorf, Australia) to separate the dsDNA. Fragment sizing was performed on AB3130 (Applied Biosystems, Australia) or AB3500 (Applied Biosystems, Australia).

Optimisation of MLVA1 method was performed to achieve the best amplification results.

MLVA2 utilised published primers from Van Cuyck *et al.* (2012). A single multiplex was performed with the four forward and reverse primers (Geneworks, Australia) (Table 6) and the protocol as outlined for MLVA1 was followed.

Table 6: MLVA2 primer sequences for genotyping *S. pneumoniae* (Van Cuyck *et al.,* 2012).

Primer	Forward	Reverse
Spneu17	F-TCGAAAATCTCTGCAAACCA	CGGACTAGGCGGCTGATTA
Spneu19	V-TCGGGTGTAGTCGTGTTTACT	AACTGATGTAGCTAAACCTAAAAAGAA
Spneu27	N-TCAGGAACAGCTATTATCCC	CCAACCTCCTTTTCGTTTCA
Spneu39	P-CCTTGGACTACCACCTCGTT	GCCGTGACAGACTTCTGGAA

The modified MLVA4 is a proposed genotyping method for *S. pneumoniae*. It also utilises BOX elements, which are amplified and separated according to fragment sizes, but alternative BOX elements already published by Koeck *et al.* (2005), Elberse *et al.* (2011a), Rakov *et al.* (2011) and Van Cuyck *et al.* (2012) were also included. MLVA4 also consisted of an alternative combination of BOX primers in master-mixes other than those already used in other MLVA methods, based on the Multiplex Manager results. The MLVA4 protocol is similar to MLVA1 except ten primers (Geneworks, Australia) and three multiplexes were used instead (Table 7). Four primers were newly re-designed.

Table 7: MLVA4 primer sequences, taken from Van Cuyck *et al.* (2012), Elberse *et al.,* (2011c) or newly designed in this study.

	VNTR	Forward primer	Reverse primer
Master Mix 1	BOX-01	F- CCAGAGACATTGATGAAGAGA	CGCTTTGATGAACTTGAGTT
	BOX-10 (new)	N- GGAGCCGAGTAGGAGATTCTCAC	TCGTAGGCTGCTACATTGACCAG
	Spneu19 (new)	V- CACTCACCGTTAGCATTGACTCG	TAATCAGGGAGTAGTTGGTTGGG
	Spneu39	P-CCTTGGACTACCACCTCGTT	GCCGTGACAGACTTCTGGAA
Master Mix 2	BOX-12 (new)	P- GAGATTGCCCTTTTCATCTTCG	AGCAACCATTGAAACGCCTG
	BOX-13 (new)	F- TCAAAAGATTGGAGAGTTCCGC	GGATTTGGAGAGCAAGCAGATC
Master Mix 3	BOX-02	N- TTGCTTGGTACAGAAAACAAA	CCCCATAAAACCCTCCTTATA
	BOX-03	V- TCCAACACGACCTTTATCC	TTCAGTAAACCCAGCTCGTA
	BOX-04	TGGGTAAAAGTAGACAGGACT	P-CACTTCTACACTAGTTTGTAAAGCA
	Spneu17	F-TCGAAAATCTCTGCAAACCA	CGGACTAGGCGGCTGATTA

3.3.6 Real-time PCR

Real-time PCR was determined to be a faster and easier method to type the capsule genes of *S. pneumoniae.* An 18µL reaction containing SybrGreen mastermix (Invitrogen, Australia), DNase and RNase free water (Invitrogen, Australia) and primers (10µM) was combined with 2µL *S. pneumoniae* DNA thermolysate in real-time PCR tubes. These were placed into a 72-well ring on a real-time PCR Qiagen RotorGene (Qiagen, Australia). Real-time PCR conditions were as follows: initialise for 10min at 52°C, cycle 40 times at 95°C for 15sec, 67°C for 20sec and 72°C for 20sec, followed by a standard melt curve (60°C – 90°C) rising 1degree/step, and finally performing a High Resolution Melt curve (60°C – 95°C) rising 0.02degrees/step.

3.4 Data analysis

All MLVA results were analysed using PeakScanner V1.0 software (Applied Biosystems, Australia). A MLVA type (MT) using MLVA1 was assigned to each isolate using the MLVA database (http://www.mlva.net) (Schouls & Van der Heide, 2012). MT types for MLVA2, and MLVA4 was manually assigned from our own database. Van Cuyck *et al.* (2012) did not provide a database or published MLVA2 types that could be used in the Queensland pneumococcal population.

The pneumococcal population structure using MLVA1, MLVA2 and MLVA4 were displayed as eBurst diagrams produced from PHYLOViZ software (Francisco *et al.,* 2012). The eBurst diagram attempts to represent the possible evolutionary relationships between strains identified by allelic profiles, including the ability to overlay information data (e.g. serotype, geographic location) on the analysis algorithm results (Francisco *et al.,* 2012). Clonal complexes (CC) are defined as two or more isolates that are genetically related according to their genotypes and linked by single locus variants (SLV) or double locus variants (DLV). The programme can expand to include triple locus variants (TLV) if required since when using MLST, a difference of three or more in the ST (out of seven) indicates a distinctly different clone (McGee *et al.,* 2001). Founding genotypes are identified by the eBurst model and is based on the genotype or ST with the most SLV links. Where an international or larger database is available, Queensland isolates were compared so as to be more confident of a founder assignment.

MLST sequences for each of the seven housekeeping genes were analysed using ChromasPro v1.6 software (Technelysium Pty Ltd.) and ClustalW batch alignment analysis. Template sequences available from the MLST database (www.mlst.net) were

used to shave down the fragment size to the correct length for submission to the MLST database. An allele number was assigned to each housekeeping gene from the MLST database based on the unique SNPs within each sequence. The combination of seven allele numbers (one for each housekeeping gene) was compiled and submitted to the MLST database to provide the sequence types (ST). The Queensland pneumococcal sequence types were displayed as an eBurst diagram using PHYLOViZ software.

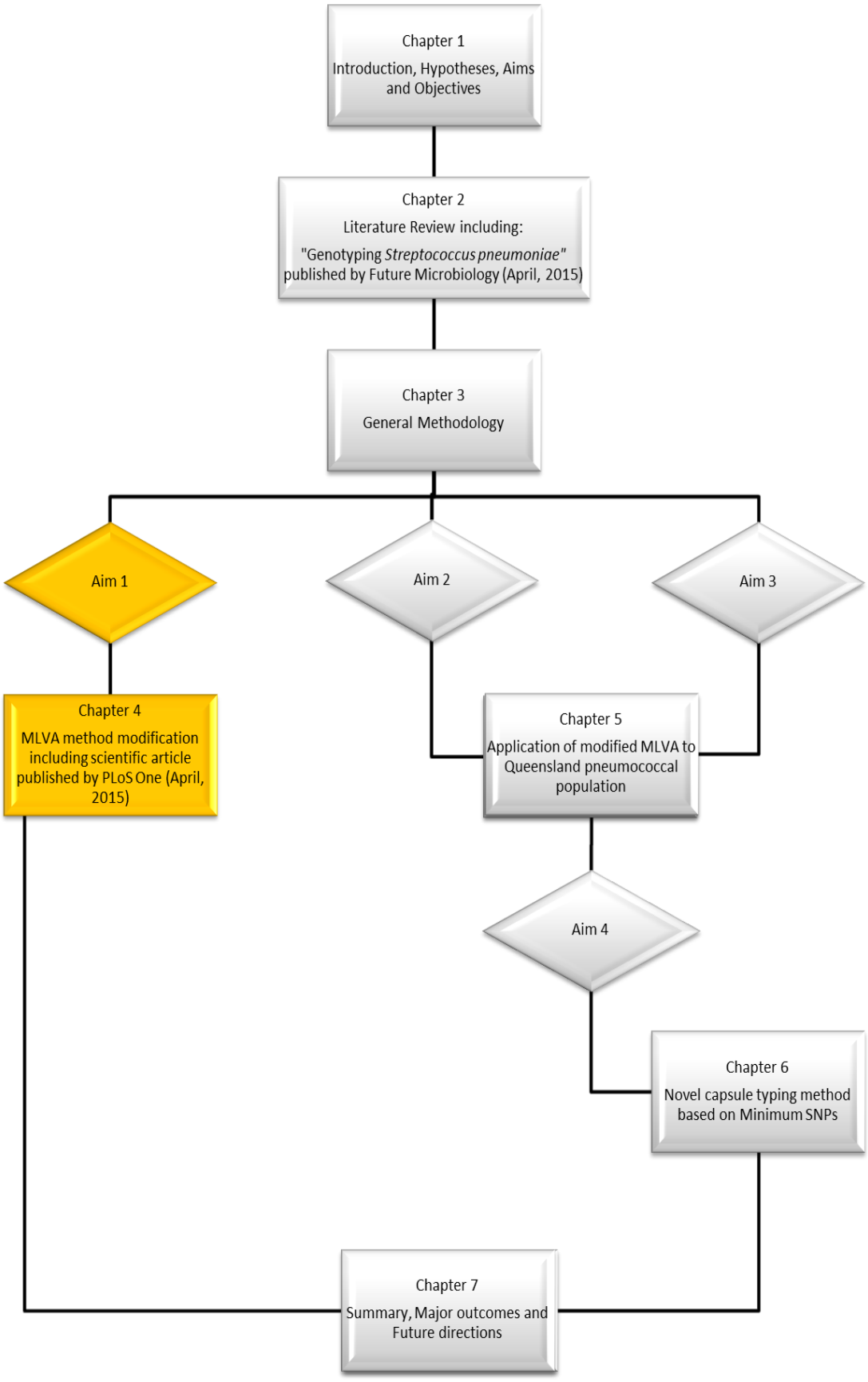
The Simpson's Index of Diversity (S) was calculated to compare the discriminatory powers of the MLVA1, MLVA2, MLVA4 and MLST methods (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool). If the 95% confidence intervals (CI) overlapped between methods, the hypothesis that the methods had similar discriminatory power could not be excluded. The Adjusted Wallace coefficient (AW) was used to measure the congruence between typing methods (Severiano *et al.,* 2011; Wallace, 1983). The Adjusted Wallace coefficient provides an "effective comparison of the results of different molecular typing studies, providing a better evaluation of the strengths and weaknesses of each study and of each typing method" (Severiano *et al.*, 2011) whereas Simpson's Index of Diversity is simply a measurement of diversity in a given population.

The frequency of non-amplified loci ('99') was compared between each MLVA method to determine whether this was a limitation of a particular method or associated with specific serotypes. Hunter-Gaston Diversity Index (DI) (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl) was used to calculate the genetic diversity of VNTR genes within the Queensland population (Platt, 2015). DI is a measure of the variation of the number of repeats at each locus, ranging from 0.0 (no diversity) to 1.0 (complete diversity) (Hunter & Gaston, 1988).

3.5 Ethics statement

Ethical clearance was not required for this study since patient information such as birth date, gender, race or location has not been examined. This epidemiology study only examines the *S. pneumoniae* isolates, changing genotypes and pneumococcal population structure within Queensland. Pneumococcal isolates were labelled with a coded number so that the original identification number could not be traced back to a particular patient.

CHAPTER 4: MODIFIED MLVA FOR GENOTYPING QUEENSLAND INVASIVE *STREPTOCOCCUS PNEUMONIAE*


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

4.1 Introduction

To combat the changing pneumococcal population, the 13vPCV was introduced in Australia in July 2011, targeting an extra six *S. pneumoniae* serotypes that were not targeted by the original 7vPCV childhood vaccine (Selvey, 2011). So far, there have been few published studies examining the pneumococcal population structure in Australia since the introduction of 13vPCV, and no studies examining Queensland isolates (Johnson *et al.,* 2012). For decades, serotyping pneumococci with antisera, such as the Quellung reaction, has been used for surveillance of serotype distribution and serotype replacement (Statens Serum Institut, 2013). However serotyping methods are expensive, laborious, sometimes ambiguous, and reveals no information about genetic recombination and capsule switching. Therefore several genotyping methods have since developed.

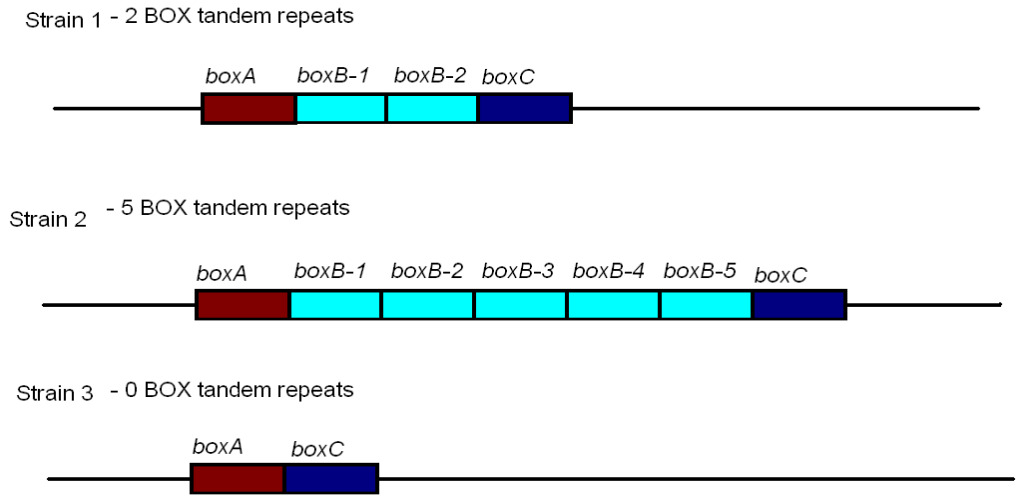
Commonly, MLST is used to genotype *S. pneumoniae* based on the original technique involving housekeeping genes developed by Enright and Spratt (1998). However, MLST is expensive and laborious, not suitable for large-scale genotyping or routine use (Elberse *et al.,* 2011a; Koeck *et al.,* 2005; Schouls & Van der Heide, 2012; Hermans *et al.,* 1995; Malachowa *et al.,* 2005). To reduce cost and labour intensity, MLVA was developed for genotyping *S. pneumoniae*. It is reported to be more discriminatory, inexpensive and faster than MLST, however more suitable for short-term epidemiology changes and localised outbreaks (Elberse *et al.,* 2011a; Koeck *et al.,* 2005; Van Cuyck *et al.,* 2012).

Different MLVA protocols exist including Koeck's *et al.* (2005) MLVA that amplifies seventeen VNTRs, each in a singleplex PCR. This technique was used to genotype pneumococcal isolates in Burkina Faso (Yaro *et al.,* 2006). Recently this protocol has been modified by Van Cuyck *et al.* (2012) by reducing the number of VNTRs to seven, a number that was experimentally determined to have the highest discrimination (Hunter-Gaston Diversity >0.8) within a population of 331 UK isolates (pre-7vPCV). This seven marker MLVA is claimed to be a minimum universal set, ideal for genotyping pneumococcal isolates. However it is known that pneumococcal populations differ between countries, therefore the selection of these seven MLVA markers may not be suitable for typing Australian isolates. These seven MLVA markers will be tested in this study.

Another MLVA protocol was developed by Elberse *et al.* (2011a), that utilises eight BOX VNTR loci amplified in two multiplex PCRs with fluorescently labelled probes. BOX fragments contain varying numbers of *boxB* repeat regions found between b*oxA* and *boxC* which remain stable under laboratory conditions (Figure 7) (Elberse *et*

al., 2011a; Martin *et al.,* 1992). Elberse's MLVA protocol has been applied to genotype pneumococci in the Netherlands, and has been used in England to track a localised outbreak of serotype 5 (Pichon *et al.* 2010). However, a limitation of this MLVA method is that approximately 25% of BOX loci fail to amplify (assigned '99') therefore leaving profiles incomplete, an issue that remains unresolved. It is unknown why BOX loci are failing to amplify but we hypothesise that these amplification failures may be of one of two reasons, 1) that primers are unable to anneal to the primer-binding site due to mutations in the *S. pneumoniae* genome, or 2) the BOX element is no longer present at all. Only sequencing of unusual size fragments have been reported (Elberse *et al.,* 2011a).

Figure 7: Three *S. pneumoniae* strains demonstrate the difference between BOX repeats for BOX-01 used in the MLVA method. *boxA* and *boxC* genes are conserved between all *S. pneumoniae* strains.



A number of theories and hypothesis for the function of BOX elements have been suggested. BOX loci are located in regions between promoters and the first downstream gene. They are presented at the 5' end of transcripts and are therefore thought to be involved in regulation of these genes (Knutsen *et al.,* 2006; Martin *et al.,* 1992). Most BOX elements have been located near virulence genes such as *neuA*, which encodes one of two neuraminidase, a putative pneumococcal virulence factor (Camara *et al.,* 1991; Martin *et al.,* 1992), and *ply*, which encodes pneumolysin (Walker *et al.,* 1987), or genes involved in gene transformation such as *comA,* which is required for induction of competence for genetic transformation (Hui & Morrison, 1991; Martin *et al.,* 1992) *hexB*, which is required for Hex mismatch repair system (Martin *et al.,* 1992;

Prudhomme *et al.,* 1989) and *mmSA* which is involved in recombination (Martin *et al.,* 1992).

BOX loci can also form stable stem-loop structures which is functionally important as it could moderate the expression of neighbouring genes (Knutsen *et al.,* 2006). Secondary structure rather than the orientation of the BOX element is thought to be more important (Knutsen *et al.,* 2006). It is also possible that BOX can stimulate the expression of downstream genes by increasing the half-life of mRNA, which is stabilised by 5' stem-loop structure (Martin *et al.,* 1992).

BOX elements have also been thought to play a role in transcription termination since approximately 3 times more BOX elements are located between genes orientated 3' to 3' than genes orientated 5' to 5' (Hoskins *et al.,* 2001). Specifically, *boxB* modules are thought to have inhibitory effects depending on the orientation, for example a forward orientated *boxB* residing at 5' end of transcript down regulates the expression of co-transcribed genes (Knutsen *et al.,* 2006). BOX elements may also enhance gene expression by stabilising mRNAs (Hoskins *et al.,* 2001). As well as this it has been investigated whether BOX loci regulate the level of expression of early competence (com) genes *qsrAB* and *comAB*, which are important for genetic transformation (Knutsen *et al.,* 2006). *ComAB* encodes and ABC transporter (*comA*) and its accessory protein *comB*. BOX elements at the 5' end transcript enhanced the expression of *qsrAB*, which encodes an ABC transporter with unknown function (Martin *et al.,* 1992). Also BOX elements were shown to have a strong positive effect on *lacZ* expression when a *lacZ* was artificially introduced in pneumococci immediately downstream of a specific promoter (Knutsen *et al.,* 2006). The authors observed effects in CSP-induced expression, however had *lacZ* been inserted elsewhere, other effects might have been seen.

Despite the uncertainty of the specific functions of these genes, BOX elements have long been recognised as a potential means for genotype classification. Hermans *et al.* (1995) determined that BOX fingerprinting (different technique to MLVA but still utilises BOX loci) was restricted to only *S. pneumoniae* since the BOX probes used in the experiment did not hybridise to 15 other Gram-positive bacteria. Likewise Van Belkum *et al.* (1996) and Koeuth *et al.* (1995) verified *boxB* typing with a single oligonucleotide primer was restricted to *S. pneumoniae* and some other streptococci. This may possibly be useful to limit the environmental contamination of other bacteria since BOX probes will not bind and amplify them. BOX typing by probe fingerprinting and by PCR fingerprinting has been claimed to be an accurate technique for typing pneumococcal strains since it provides a high degree of discrimination (Hermans *et al.,* 1995; Van

Belkum *et al.,* 1996). Hermans *et al.* (1995) demonstrated that BOX fingerprinting could be used to classify 21 types out of 28 isolates, and shows that it had a higher discriminatory power than PFGE, ribotyping, PCR fingerprinting and serotyping. Despite using a different method, this does indicate that the use of BOX loci can be highly discriminatory. BOX-PCR is less reliable as it produces a multi-band pattern of "added polymorphisms of multiple BOX elements" (Koeck *et al.,* 2005). MLVA however analyses each BOX locus separately (Koeck *et al.,* 2005).

Finally, Rakov *et al.* (2011) developed Multi-Locus *boxB* Typing (MLBT), a variation of MLVA by sequencing VNTR loci to detect Single Nucleotide Polymorphisms (SNPs) as well as fragment length variations. MLBT contains VNTRs that have been used in the other MLVA protocols, however the complexities of MLBT does not enable ease of comparison with other MLVA methods (Elberse *et al.,* 2011a; Koeck *et al.,* 2005; Rakov *et al.,* 2011). These complexities come out of the fact that VNTRs are sequenced, and pneumococcal profiles are based on SNPs within these sequences.

The aim of the present study was to generate a fast, inexpensive and highly discriminatory MLVA method (hereby called MLVA4) for genotyping invasive *S. pneumoniae* in Queensland. The objective of developing MLVA4 was to ensure that it would provide a more accurate reflection of the *S. pneumoniae* population structure compared to MLST and previously published MLVA methods. As there is no nationally accepted MLVA technique for *S. pneumoniae*, previously published MLVA methods were used for comparison, as well as MLST.

4.2 Methods

4.2.1 Laboratory methods

MLVA markers were selected from Elberse *et al.* (2011a) (BOX-01, BOX-02, BOX-03, BOX-04a, BOX-04b, BOX-06a, BOX-06b, BOX-11, BOX-12 and BOX-13), Koeck *et al.* (2005) (Spneu17, Spneu19, Spneu27, Spneu31, Spneu39), Van Cuyck *et al.* (2012) and Rakov *et al.* (2011) (B10). In this study, four primer sets were redesigned due to difficulties in amplification, including BOX-12 (F: 5'-GAGATTGCCCTTTTCATCTTCG-3'; R: 5'-AGCAACCATTGAAACGCCTG-3'), BOX-13 (F: 5'-TCAAAAGATTGGAGAGTTCCGC-3'; R: 5'-GGATTTGGAGAGCAAGCAGATC-3'), Spneu19 (F: 5'-CACTCACCGTTAGCATTGACTCG-3'; R: 5'-TAATCAGGGAGTAGTTGGTTGGG-3') and B10 (F: 5'-GGAGCCGAGTAGGAGATTCTCAC-3'; R: 5'-TCGTAGGCTGCTACATTGACCAG-3') (Geneworks, Australia). All MLVA markers were analysed *in silico* using nBlast and the Tandem Repeat Finder algorithm to determine the expected fragment sizes and copy

numbers against five known *S. pneumoniae* genomes (R6, G54, CGSP14, TIGR4 and Hungary19A-6) from the NCBI database.

Elberse's MLVA1 contained two multiplexes with eight BOX elements (Elberse *et al.,* 2011a). The MLVA2 procedure was based on Van Cuyck's MLVA method and contained seven VNTRs (Spneu17, Spneu19, Spneu25, Spneu27, Spneu33, Spneu37 and Spneu39). A single multiplex reaction was performed with Spneu17, Spneu19, Spneu27 and Spneu39 as the other three had already been typed in MLVA1. Spneu31 and B10 were separately amplified to determine whether they provided high discrimination within the pneumococcal population, and therefore suitable for the modified MLVA method.

Hence, MLVA4 was the modified MLVA method based on the discrimination of the fourteen loci previously used, and the ability to amplify all loci across 35 different serotypes detected in Queensland. Three multiplexes were developed with primers BOX-01, BOX-02, BOX-03, BOX-04, B10, BOX-12, BOX-13, Spneu17, Spneu19 and Spneu39. The loci used in each multiplex reaction are presented for each of the MLVA typing methods (Table 8).

Table 8: The Multiplex arrangement of VNTR loci in each MLVA method used in this study.

Genotyping method	Multiplex #	Loci	Reference
MLVA1	1	BOX-01, BOX-02, BOX-03, BOX-04	Elberse *et al.,* 2011a
	2	BOX-06, BOX-11, BOX-12, BOX-13	
MLVA2	1	Spneu17, Spneu19, Spneu27, Spneu39	Van Cuyck *et al.,* 2012
	2	Spneu25, Spneu33, Spneu37	
MLVA4	1	BOX-01, B10, Spneu19, Spneu39	This study
	2	BOX-12, BOX-13	
	3	BOX-02, BOX-03, BOX-04, Spneu17	

MLVA genotyping was performed as described in Chapter 3. The MLVA4 method initially was applied to three *S. pneumoniae* control strains with known MLVA profiles for all loci. A further 170 isolates, covering all 35 serotypes detected in Queensland from 2007-2012 were genotyped using MLVA4 for validation, before a further 144 isolates (total of 317 isolates) were also genotyped using MLVA4. The MLVA4 method has novel aspects as four sets of primers and the three multiplex PCR has been newly designed in this study.

Finally, MLST was also applied to 202 selected isolates for comparison purposes using the Corbett Cas1200 Robotics (Qiagen, Germany) for automation (Enright & Spratt, 1998; Jefferies *et al.,* 2003). Isolates selected for MLST included 13vPCV serotypes (except serotype 19A) and non-13vPCV serotypes to minimise labour and

costs, but also to allow examination of changes in these recently targeted or non-targeted serotypes.

4.2.2 Analysis

MLVA results were analysed using PeakScanner V1.0 software (Applied Biosystems, Australia). A MLVA type (MT) was assigned to each pneumococcal isolate depending on the MLVA protocol – refer to Chapter 3. The pneumococcal population structure using MLVA1, MLVA2 and MLVA4 were displayed as eBurst diagrams produced from PHYLOViZ software (Francisco *et al.,* 2012).

MLST results were analysed using ChromasPro software (Technelysium Pty Ltd.) using batch alignment analysis. Allele numbers and sequence types (ST) were assigned to each isolate from the MLST database (www.mlst.net) and displayed as an eBurst diagram using PHYLOViZ software.

Simpson's Index of Diversity (S) was calculated to compare the pneumococcal population discrimination when using MLVA1, MLVA2, MLVA4 and MLST methods (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool). The Adjusted Wallace coefficient (AW) was used to measure the congruence between typing methods (Severiano *et al.,* 2011; Wallace, 1983).

The frequency of non-amplified loci ('99') was compared between each MLVA method to determine whether this was a limitation of a particular method or associated with specific serotypes. Hunter-Gaston Diversity Index (DI) (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl) was used to calculate the genetic diversity of VNTR genes within the Queensland population (Hunter & Gaston, 1988).

4.3 Results

All 317 *S. pneumoniae* Queensland isolates in this study were assigned an MLVA genotype, covering all 35 serotypes detected in Queensland from 2007-2012. The MLVA4 method has novel aspects as four sets of primers and the three multiplex PCRs have been newly designed in this study. Isolates selected for MLST comparison included 13vPCV serotypes (excluding those originally in the 7vPCV i.e. serotype 4, 6B, 9V, 14, 18C, 19F and 23F, and excluding serotype 19A) and non-13vPCV serotypes to minimise labour and costs. Studies have already shown that a number of 7vPCV serotypes have significantly declined, so this study focused on the examination of these recently targeted or non-targeted serotypes.

In early experimental stages, amplification of MLST sequences and MLVA1 fragments were verified using gel electrophoresis (Figure 8 and Figure 9, respectively). Strong amplification of all seven housekeeping genes is observed for MLST. When observing the MLVA multiplexes, some isolates only have three bands observed rather than four, however it is possible that some bands have similar fragment size and therefore cannot be distinguished using gel electrophoresis. This can be observed in Figure 10.

Figure 8: Quantitative gel electrophoresis and amplification check of MLST genes in S. pneumoniae.

The Low Marker DNA Ladder (LMDL) is used as a comparison to determine the quantity of amplified housekeeping genes. The seven housekeeping genes are shown to successfully amplify for *S. pneumoniae* isolates (ID for *S. pneumoniae* isolate is above each lane). *S. pneumoniae* control 1, C1; *S. pneumoniae* control 2, C2.

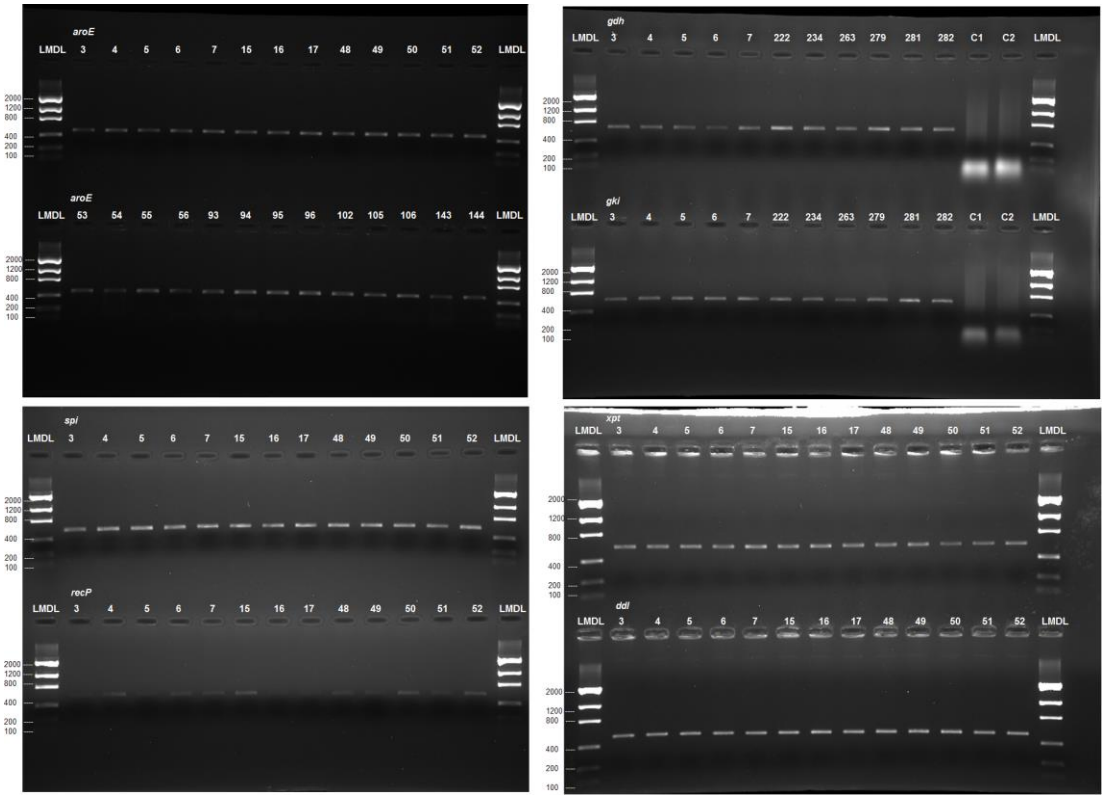


Figure 9: Gel electrophoresis of MLVA1 – multiplex 2 applied to *S. pneumoniae* isolates.

Multiplex 2 contains BOX-06, BOX-10, BOX-11 and BOX-12. *S. pneumoniae* isolates ID 1-13 have been used to show the amplification of the four BOX loci. A negative control is included in lane 14 and a DNA ladder 123bp in the final lane. Negative control, -ve. Positive control 1 (serotype 19F).

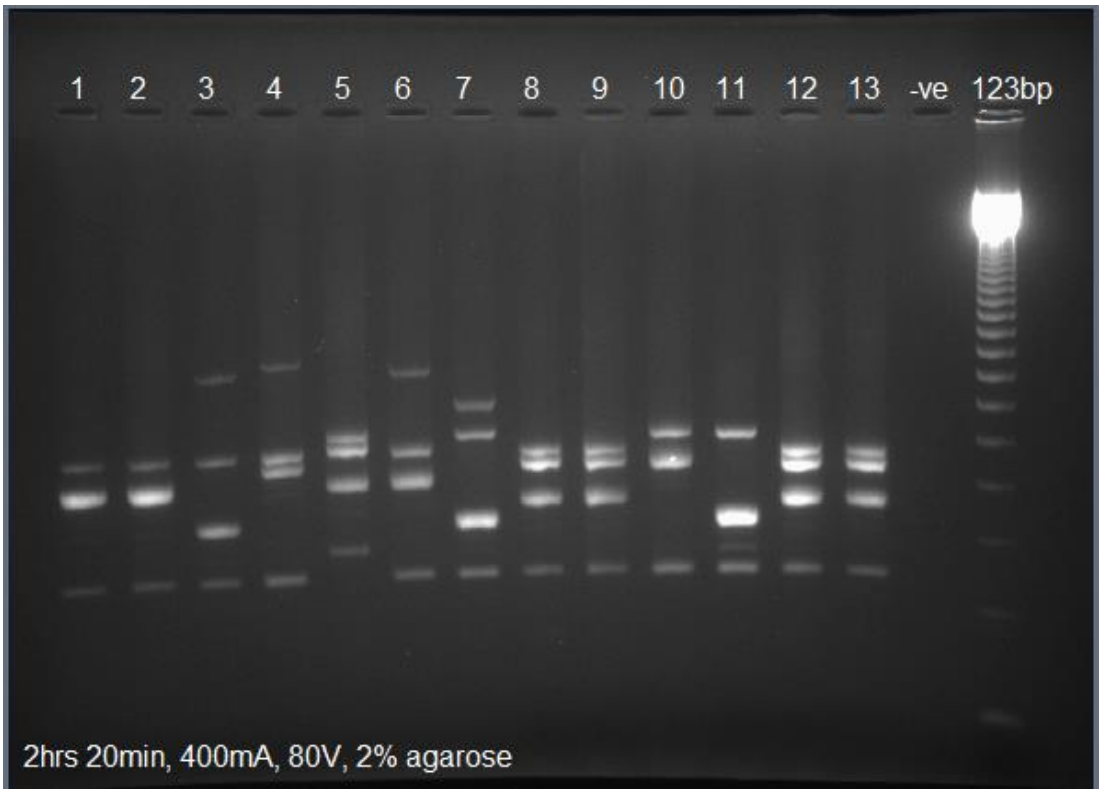
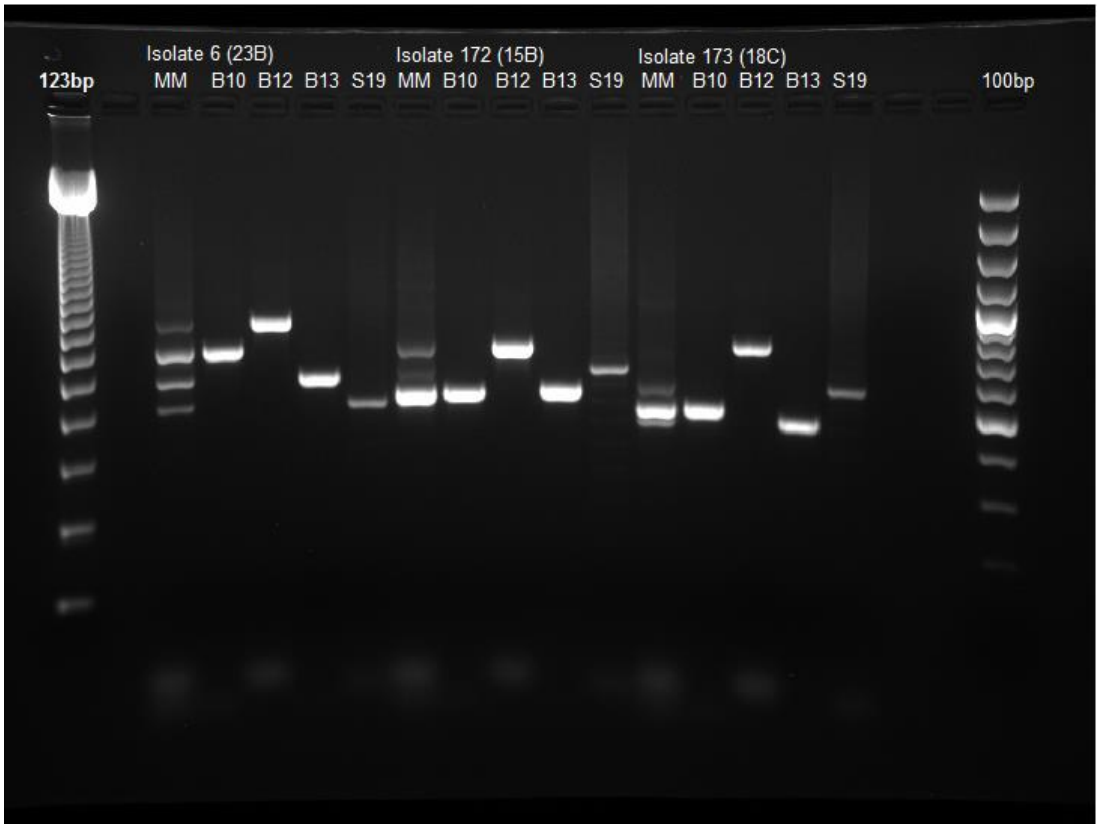


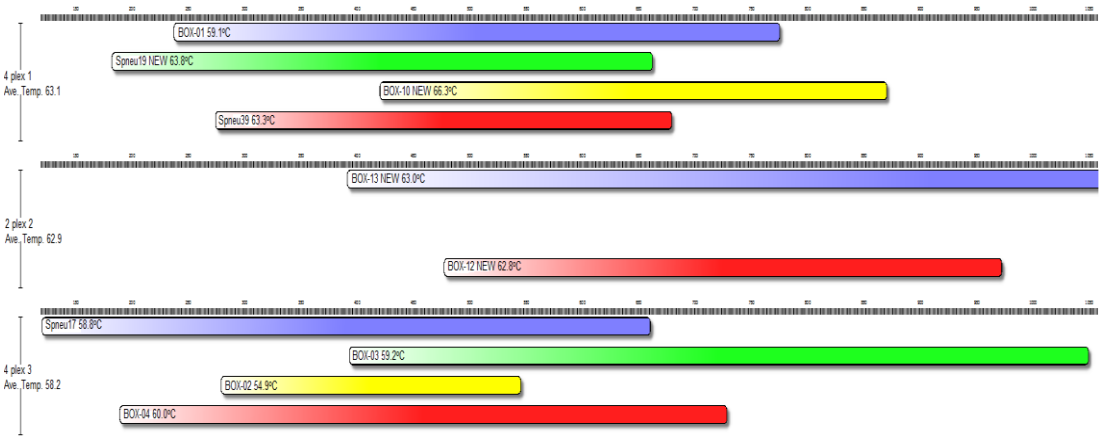
Figure 10: MLVA4 applied to three *S. pneumoniae* isolates.

S. pneumoniae isolates are listed above each grouping with respective serotype in brackets. Master Mix (MM) lane contains all four loci (BOX-10, BOX-12, BOX-13 and Spneu19) and the respective BOX-10 (B10), BOX-12 (B12), BOX-13 (B13) and Spneu19 (S19) singleplex in following lanes. A DNA marker ladder 123bp and 100bp is present in lane 1 and lane 20, respectively. It is unknown why B12 for isolate 173 is not appearing in the MM, although it is hypothesised that it could be due to faint amplification or primer dimers in MM.



Furthermore, Multiplex Manager was used to determine the best combination of MLVA primer in the minimum number of multiplexes. The results indicated that three multiplex would be optimal for amplifying all 10 VNTRs with the minimum number of reactions (Figure 11).

Figure 11: Multiplex Manager shows the grouping of 10 VNTRs into three multiplexes for MLVA4.



Using Microsoft Excel software, a database was generated containing all the MLVA fragment sizes as analysed by PeakScanner. The fragment sizes were interpreted into tandem repeats (Table 9) and the combination of fragment sizes provided the allele code for each fragment (Figure 12).

Table 9: Size alleles for each BOX locus, assigned according to the fragment size that is identified by the DNA sequencer for MLVA4.

BOX locus	Allele	Size	BOX locus	Allele	Size	
BOX-01	0	237.0	BOX-12 (new)	0	472.0	NEW
45bp consensus size	1	280.0	45bp consensus size	1	517.0	
	2	325.0		2	562.0	
	3	370.0		3	607.0	
	4	415.0		4	652.0	
	5	460.0		5	697.0	
	6	505.0		6	742.0	
	7	550.0		7	787.0	
	8	595.0		8	833.0	
	9	639.7		9	877.0	
	10	685.6		10	922.0	
	11	731.9		11	967.0	
	12	775.3		12	1012.0	
BOX-02	0	279.0	BOX-13 (new)	0	389.0	
45bp consensus size	1_01	286.0	45bp consensus size	1	434.0	
	1	322.5		2	479.0	
	2_01	330.0		3	524.0	
	2_02	359.0		4	569.0	
	2	366.0		5	614.0	
	3	411.0		6	659.0	
	4	454.5		7	704.0	
	5	499.0		8	749.0	
	6	545.2		9	794.0	
BOX-03	1	393.0		10	839.0	
45bp consensus size	2	435.4		11	884.0	

BOX locus	Allele	Size	
	3	480.0	
	4_01	490.0	
	4	525.0	
	5	570.0	
	5_01	578.0	
	6	614.3	
	7	657.0	
	8	702.0	NEW
	8_01	732.0	
	9	747.0	
	9_01	753.0	
	10	789.0	
	11	834.0	
	12_01	860.7	
	12	879.0	
	13	918.0	
	14	963.0	
	15	1003.2	
	16	1048.8	
BOX-04	0	189.0	
45bp consensus size	1	234.9	
	2	279.5	
	3	325.6	
	4	370.1	
	5	415.2	
	6_01	453.8	
	6	459.3	
	7	504.6	
	8	548.3	
	9	595.3	
	10	638.3	NEW
	11	683.3	NEW
	12	728.3	NEW
BOX-10 (new)	0	420.0	
45bp consensus size	1	465.0	
	2	510.0	
	3	555.0	
	4	600.0	
	5	645.0	
	6	690.0	
	7	735.0	
	8	780.0	
	9	825.0	
	10	870.0	
	11	915.0	
	12	960.0	

BOX locus	Allele	Size
	12	929.0
Spneu17	0	75.0
45bp consensus size	1	120.0
	2	165.0
	3	210.0
	4	255.0
	5	300.0
	6	345.0
	7	390.0
	8	435.0
	9	480.0
	10	525.0
	11	570.0
	12	615.0
	13	660.0
	14	705.0
	15	750.0
Spneu19 (new)	0	122.0
60bp consensus size	1	182.0
	2	242.0
	3	302.0
	4	362.0
	5	422.0
	6	482.0
	7	542.0
	8	602.0
	9	662.0
	10	722.0
	11	782.0
	12	842.0
BOX-39	0	222.0
45bp consensus size	1	267.0
	2	312.0
	3	357.0
	4	402.0
	5	447.0
	6	492.0
	7	537.0
	8	582.0
	9	627.0
	10	672.0
	11	717.0
	12	762.0

Figure 12: Example of MLVA1 fragment sizing as recorded in Microsoft Excel program.

The isolate ID is presented as QRRJS # followed by the serotype below. Multiplex 1 and 2 represent the two multiplexes used in MLVA1. The fluorescent markers are colour-coded for each BOX locus (e.g. FAM is blue for BOX-01 and BOX-13). 'Size' of the fragment (bp) is recorded from the DNA sequencer, followed by the 'height' of each fluorescent peak. The 'MLVA size' is determined from the MLVA database, which is the closest to the fragment 'size' recorded. From this the 'allelic profile' is assigned to each fragment size based from the MLVA database.

	MLVA Protocol								
1	Isolate	QRRJS 1							
	Serotype	19F							
	Multiplex				1				2
		Size	Height	MLVA size	Allelic profile	Size	Height	MLVA size	Allelic profile
	FAM (1, 13)	0	0		**99**	473.9	8292	476.1	**4**
	NED (2, 6)	365.1	8740	366	**2**	284	1740	289	**2**
	VIC (3, 11)	611	8849	614.3	**6**	548.2	3227	549.7	**1**
	PET (4, 12)	726.4	1050	728.3	**12**	469.8	7179	471	**0**
88	Isolate	QRRJS 88							
	Serotype	19A							
	Multiplex				1				2
		Size	Height	MLVA size	Allelic profile	Size	Height	MLVA size	Allelic profile
	FAM (1, 13)	458.3	8322	460	**5**	475	8293	476.1	**4**
	NED (2, 6)	410.3	8636	411	**3**	284	1261	289	**2**
	VIC (3, 11)	523.3	8978	525	**4**	547.3	3985	549.7	**1**
	PET (4, 12)	367.7	7510	370.1	**4**	559.6	2386	562.3	**2**
89	Isolate	QRRJS 89							
	Serotype	19A							
	Multiplex				1				2
		Size	Height	MLVA size	Allelic profile	Size	Height	MLVA size	Allelic profile
	FAM (1, 13)	458.3	2306	460	**5**	475.3	8231	476.1	**4**
	NED (2, 6)	410	4517	411	**3**	283.9	2151	289	**2**
	VIC (3, 11)	523.3	3142	525	**4**	547.2	7553	549.7	**1**
	PET (4, 12)	367.6	3692	370.1	**4**	559.6	4396	562.3	**1**
90	Isolate	QRRJS 90							
	Serotype	1							
	Multiplex				1				2
		Size	Height	MLVA size	Allelic profile	Size	Height	MLVA size	Allelic profile
	FAM (1, 13)	324	3259	325	**2**	518.1	4142	521.3	**5**
	NED (2, 6)	409.7	1822	411	**3**	335.7	1884	334	**3**
	VIC (3, 11)	435.6	2765	435.4	**2**	547.3	2775	549.7	**1**
	PET (4, 12)	367.5	825	370.1	**4**	782.6	134	784.1	**7**

All three MLVA genotyping methods had a higher discriminatory power compared to MLST when examining Simpson's Index of Diversity (s) (Table 10). MLVA4 had the highest discrimination of S=0.990 with 203 MLVA types (n=317) when compared to the other two MLVA techniques (MLVA1 S=0.984, and MLVA2 S=0.987); however given that the 95% CI overlap between MLVA1, MLVA2 and MLVA4, the Simpson Index of Diversity is not significantly different, or in other words, the diversity is comparable. MLVA4 also had a significantly higher discriminatory power S=0.978 with 106 MLVA types (n=202) when compared to MLST method (n=202). In comparison, MLVA1 had a discrimination of S=0.963 (n=202; 71 MLVA types) and MLVA2 with S=0.977 (n=202; 97 MLVA types), both comparable to MLVA4 as 95% CI

overlap between all three MLVA methods. MLST had the lowest discrimination of S=0.936 and 49 STs (n=202). When examining MLVA1, 20.2% of isolates still contain at least one non-amplified locus. Even MLVA2 contained 21.1% of isolates with non-amplified loci. MLVA4 method reduces the '99' to 12.6%.

Table 10: Calculation of Simpson's Index of Diversity for serotyping (Quellung), MLST, MLVA1, MLVA2 and MLVA4 methods.

Total Number of *S. pneumoniae* isolates = 202						
Molecular method	**Total number of genotypes**	**Simpson's Index of Diversity (S)**	**CI (95%)**	**Average time to genotype 48 isolates (days)**	**Cost per isolate (PCR to genotype) (AUS$)**	**Non-amplified loci (%)**
Serotyping	29	0.912	0.893-0.930	-	-	N/A
MLST	49	0.936	0.920-0.952	16-20	346.65	N/A
MLVA1	71	0.963	0.953-0.974	3-4	17.10	20.8
MLVA2	97	0.977	0.970-0.985	3-4	16.37	24.3
MLVA4	106	0.978	0.971-0.986	3-4	23.71	12.4
Total Number of *S. pneumoniae* isolates = 317						
Serotyping	35	0.914	0.897-0.932	-	-	N/A
MLVA1	163	0.984	0.980-0.989	3-4	17.10	20.2
MLVA2	175	0.987	0.983-0.991	3-4	16.37	21.1
MLVA4	203	0.990	0.987-0.994	3-4	23.71	12.6

However, when comparing the Adjusted Wallace Coefficient of MLVA4 with the other two MLVA methods, all MLVA methods had similar discriminatory powers (Table 11). It indicates that MLVA4 has high congruence with MLVA1 (AW = 0.883), MLVA2 (AW = 0.766) and MLST (AW = 0.966). This indicates that the MLVA4 genotypes will have a 96.6% probability that it will have the same MLST type. Conversely, the congruency of MLST with MLVA4 (AW = 0.314) indicates that the MLST types will have a 31.4% probability of having the same MLVA4 type, indicating that MLVA4 is more discriminatory.

Table 11: Adjusted Wallace Coefficient with 95% confidence intervals (CI) for four genotyping and one serotyping method applied to invasive *S. pneumoniae*.

Total number of *S. pneumoniae* isolates = 202					
Typing method	MLVA1	MLVA2	MLVA4	Serotyping	MLST
MLVA1		0.524 (0.461-0.5287)	0.512 (0.442-0.582)	0.940 (0.891-0.989)	0.984 (0.974-0.995)
MLVA2	0.871 (0.800-0.942)		0.739 (0.667-0.810)	0.880 (0.804-0.956)	0.916 (0.887-0.945)
MLVA4	0.883 (0.823-0.942)	0.766 (0.690-0.843)		0.918 (0.842-0.994)	0.966 (0.927-1.000)
Serotyping	0.372 (0.301-0.442)	0.209 (0.157-0.262)	0.211 (0.161-0.260)		0.648 (0.587-0.709)
MLST	0.552 (0.456-0.648)	0.309 (0.244-0.374)	0.314 (0.254-0.374)	0.919 (0.869-0.969)	

New MLVA1 BOX alleles were also found with varying numbers of repeats including BOX-04 fragment sizes with 10 repeats (n=1), 11 repeats (n=1) and 12 repeats (n=3), BOX-03 fragment size 8 repeats (n=2), BOX-12 fragment size 0 repeats (n=2) and BOX-13 fragment size >2000bp (unsure how many repeats until sequenced; n=8). There are 107 new MLVA1 type (MT) types not recorded by the MLVA database (Schouls & Van der Heide, 2012). The BOX lengths for MLVA4 were previously calculated (Table 9). A total of 219 MLST types were submitted to the MLST international database (id # 23211 – 23429) of which 19 isolates had new MLST types (Table 12).

Table 12: New pneumococcal MLST genotypes identified in Queensland from 2007 to 2012.

The MLST genotypes were submitted to the MLST international database for assignation of the new sequence types. Sequence type, ST; Australia, AUS; Queensland, QLD.

ST	aroE	gdh	gki	recP	spi	xpt	ddl	Strain	Country	Capsule type	Region	Year of isolation
8969	2	13	4	1	9	19	14	QRRJS20	AUS	6C	QLD	2007
9673	26	1	2	14	9	16	19	QRRJS275	AUS	3	QLD	2007
9662	2	40	17	19	10	1	27	QRRJS281	AUS	10A	QLD	2007
9671	25	13	4	16	6	1	17	QRRJS283	AUS	10A	QLD	2007
9665	5	43	6	1	9	127	18	QRRJS220	AUS	34	QLD	2008
9670	18	12	4	44	9	77	97	QRRJS221	AUS	35B	QLD	2008
9672	25	31	17	16	32	1	44	QRRJS263	AUS	18A	QLD	2008
9664	5	5	9	9	6	1	19	QRRJS144	AUS	23B	QLD	2009
9661	2	8	62	16	6	19	309	QRRJS170	AUS	10F	QLD	2009
9673	26	1	2	14	9	16	19	QRRJS11	AUS	3	QLD	2010
9663	2	264	29	18	42	395	18	QRRJS143	AUS	22F	QLD	2011
9666	7	41	47	16	6	1	44	QRRJS52	AUS	15C	QLD	2011
9664	5	5	9	9	6	1	19	QRRJS6	AUS	23B	QLD	2011
9082	25	31	4	16	32	1	44	QRRJS96	AUS	18A	QLD	2011
9082	25	31	4	16	32	1	44	QRRJS290	AUS	18A	QLD	2012
9668	10	11	10	1	6	8	1	QRRJS335	AUS	11A	QLD	2012
9669	10	41	2	6	10	3	1	QRRJS336	AUS	15B	QLD	2012
9660	1	8	9	2	6	20	6	QRRJS363	AUS	23B	QLD	2012
9667	8	5	27	5	1	1	1	QRRJS375	AUS	16F	QLD	2013

4.3.1 VNTR loci and '99'

A Hunter-Gaston Diversity Index (DI) indicated that BOX-01, BOX-04, BOX-12, BOX-13 and Spneu17 had the highest diversity (DI ≥ 0.80) while BOX-02, BOX-11, Spneu19 and Spneu27 had the lowest diversity (DI ≤ 0.66) (Table 13). BOX-02 had low discriminatory power and was used in MLVA4 as an anchor marker. Spneu19 was used in MLVA4 because it had highest discrimination for specific serotypes such as serotype 8, 18A and 18C (Table 14). A graphical representation of the different alleles for each locus is provided in *Appendix A3: Graphical representation of the diversity of MLVA genes in* S. pneumoniae. In this study '99' was assigned to VNTR loci that failed to amplify. In particular, BOX-06, BOX-13 and Spneu19 resulted in the most '99' even after singleplex PCR. Singleplex PCR resolved >70% of BOX-02, BOX-03 and BOX-12 that previously didn't amplify. Singleplex PCR follows the same protocol as multiplex PCR, however only one VNTR locus is amplified per reaction.

Table 13: The Hunter-Gaston Diversity for all selected MLVA loci used in this study (n=317 *S. pneumoniae* isolates) and frequency of non-amplified loci '99'.

The respective MLVA method that the loci were included in is shown followed by the Hunter-Gaston Diversity (DI) and the 95% confidence interval (CI). The number of alleles for each locus (K) is listed. The number of non-amplified loci is shown before and after singleplex amplification, and the identification of specific serotypes that contain non-amplified loci even after singleplexing is shown.

MLVA method locus is used in	Locus	DI	CI	Size of tandem repeat (bp)	K	# non-amplified loci before singleplex	# non-amplified loci after singleplex	Non-amplified serotypes (#)
MLVA2, MLVA4	Spneu17	0.853	0.840 – 0.866	45	12	6	0	-
MLVA1, MLVA2, MLVA4	BOX-13	0.821	0.806 – 0.836	45	10	23	13	33F (5); 6C (7); 19F (1)
MLVA1, MLVA2, MLVA4	BOX-12	0.805	0.778 – 0.833	45	13	89	4	3 (3); 23B (1)
MLVA1, MLVA4	BOX-01	0.801	0.785 – 0.817	45	10	11	4	19F (3), 33B (1)
MLVA1, MLVA4	BOX-04	0.797	0.771 – 0.823	45	12	7	0	-
MLVA2, MLVA4	Spneu39	0.794	0.775 – 0.812	45	10	0	0	-
MLVA1, MLVA4	BOX-03	0.789	0.754 – 0.824	45	14	28	0	-
MLVA4	BOX-10	0.784	0.748 – 0.821	45	11	54	4	15C (2); 19A (2)
	Spneu31	0.729	0.696 – 0.762	45	11	7	2	14 (1); 18A (1)
MLVA1, MLVA2	BOX-06	0.696	0.653 – 0.739	45	8	53	31	7F (26); 19A (4); 19F (1)
MLVA2, MLVA4	Spneu19	0.663	0.632 – 0.694	60	9	24	16	3 (12); 38 (4)
MLVA1, MLVA4	BOX-02	0.651	0.628 – 0.674	45	5	1	0	-
MLVA2	Spneu27	0.637	0.601 – 0.674	45	8	13	3	18A (1); 22A (1); 23B (1)
MLVA1	BOX-11	0.392	0.341 – 0.443	45	3	14	7	19A (2); 6A (2); 6C (1): 15C (2)
	TOTAL	-	-	-	-	**330**	**121**	

For further interest, the number of VNTRs was collated to determine which VNTR loci had the highest discrimination within each of the 30 serotypes found in Queensland (Table 14). This was used to aid in the selection of highly discriminatory loci for the MLVA4 method, based on the discriminatory value within each serotype. Some pneumococcal serotypes were few in number in the Queensland population; therefore there may be more variance than detected here.

Table 14: Assignation of the VNTR allele for each *S. pneumoniae* serotype from this study.

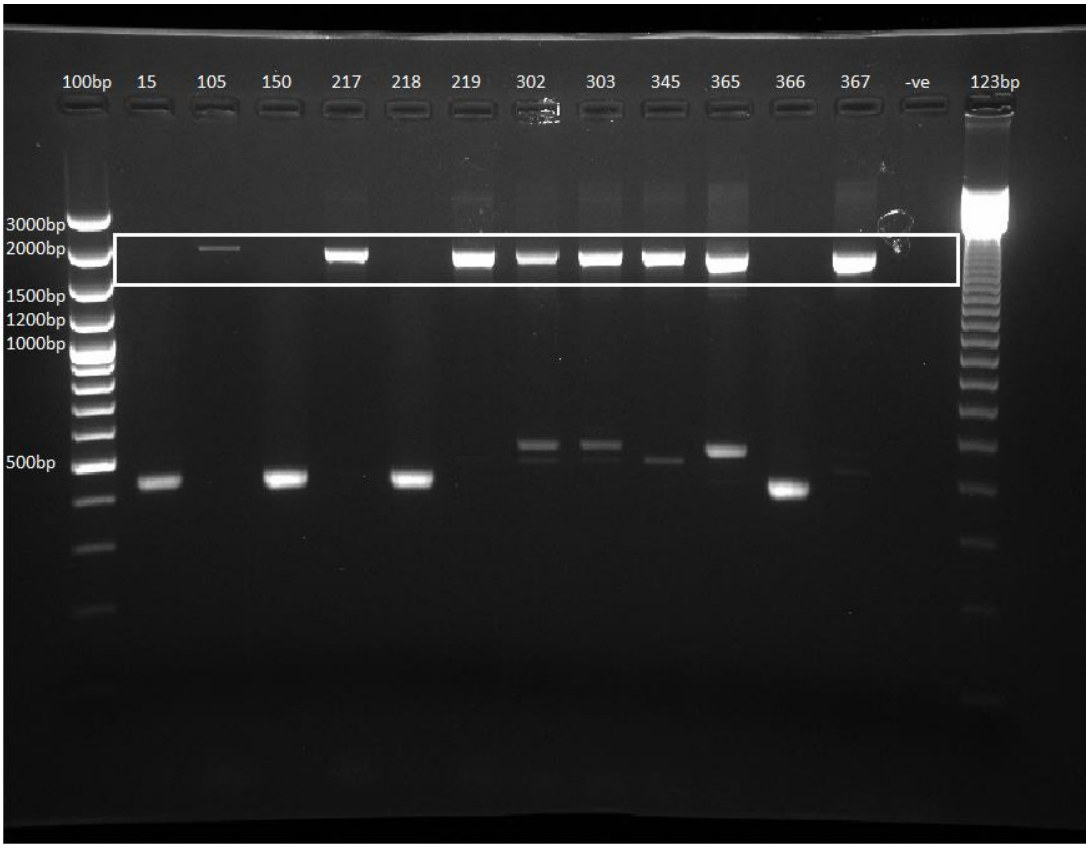
A total of 30 serotypes that were detected in Queensland are presented, followed by the number of alleles for each VNTR locus. For example, this study detected 4 alleles of BOX-01 in serotype 3, but only 1 allele in serotype 1. Yellow highlighted numbers indicate the BOX loci with the highest observed number of polymorphisms for each serotype, e.g. BOX06 and BOX13 had the highest observable number of polymorphisms (4 polymorphisms) for *S. pneumoniae* serotype 1.

	VNTR Loci used in any MLVA method													
Serotype	BOX01	BOX 02	BOX 03	BOX 04	BOX 06	BOX 10	BOX 11	BOX 12	BOX 13	Spn 17	Spn 19	Spn27	Spn31	Spn39
1	1	2	2	1	4	2	2	3	4	2	2	1	2	2
3	4	2	1	2	1	4	2	2	4	6	3	1	3	2
4	4	3	2	3	2	2	1	3	3	2	3	2	2	2
6A	4	2	5	3	2	3	2	3	5	3	2	2	2	5
6B	2	1	2	2	2	2	1	2	2	1	2	2	2	2
6C	2	2	5	4	2	4	3	5	5	6	3	3	5	2
7F	2	2	2	1	4	7	1	1	1	3	1	1	1	2
8	1	1	1	2	1	1	1	1	1	1	3	1	1	1
9N	2	1	2	1	1	3	1	2	1	1	2	2	1	1
9V	1	2	1	2	2	3	1	3	1	2	1	1	2	2
10A	2	2	2	3	2	1	2	2	2	2	2	2	2	3
10F	1	1	1	1	1	2	1	1	1	1	1	1	1	1
11A	2	2	2	2	2	2	1	2	2	2	1	2	2	2
12F	1	1	1	1	1	1	1	1	1	1	1	2	1	1
14	3	2	1	2	2	1	2	3	1	1	2	2	3	1
15B	2	2	2	5	1	3	2	4	2	1	2	2	1	1
15C	3	3	2	4	2	3	3	2	3	3	2	2	2	3
18A	3	2	1	1	1	2	1	1	1	1	3	2	2	1
18C	1	1	2	1	2	1	1	3	2	1	4	1	1	3
19A	6	3	8	6	6	7	3	10	5	6	5	7	8	5
19F	5	5	7	7	3	6	2	6	5	7	4	4	8	5
22F	1	1	3	3	2	1	2	2	2	2	1	2	1	2
23A	1	1	1	3	1	2	1	1	1	1	2	1	2	3
23B	5	3	4	3	2	3	2	6	3	4	4	3	1	3
23F	3	3	3	2	2	4	2	4	2	3	2	1	3	4
33F	2	2	1	3	2	4	1	2	3	2	3	7	2	2
34	1	1	2	2	1	1	2	2	2	1	1	1	2	1
35B	4	2	3	2	1	1	2	1	1	2	2	1	2	2
35F	1	1	3	1	1	1	1	1	1	2	2	2	2	2
38	2	1	2	3	1	2	1	2	1	1	1	1	1	1

In total, MLVA1 contained 20.8% isolates with at least one non-amplified gene, MLVA2 had 24.3% non-amplified and MLVA4 had the lowest at 12.4% non-amplified. These VNTR loci failed to amplify in specific serotypes. BOX-06 failed to amplify in 75% of serotype 7F (27 isolates), BOX-13 failed to amplify in 42% of serotype 33F (5 isolates) and in 54% of serotype 6C (7 isolates), Spneu19 failed to amplify in 63% of serotype 3 (12 isolates) and in 50% of serotype 38 (4 isolates), and BOX-01 failed to amplify in 20% of serotype 19F isolates (3 isolates) (Table 13). Spneu19 allele sequences were not found in serotype 3 genomes using BlastN (strains OXC141, SP3, SPN021198, SPN034156, SPN034183, SPN072838, SPN994038 and SPN994039). Similarly Spneu19 failed to amplify in serotype 38; however no serotype 38 genome has been sequenced to date to allow us to determine whether the Spneu19 locus exists. No match was found for BOX-13 in serotype 6C NCBI genomes, and failed to amplify in serotype 33F, however gel electrophoresis showed a large >2000bp fragment (Figure 13).

Figure 13: Amplification of >2000bp fragments of BOX-13 in *S. pneumoniae* serotype 33F isolates.

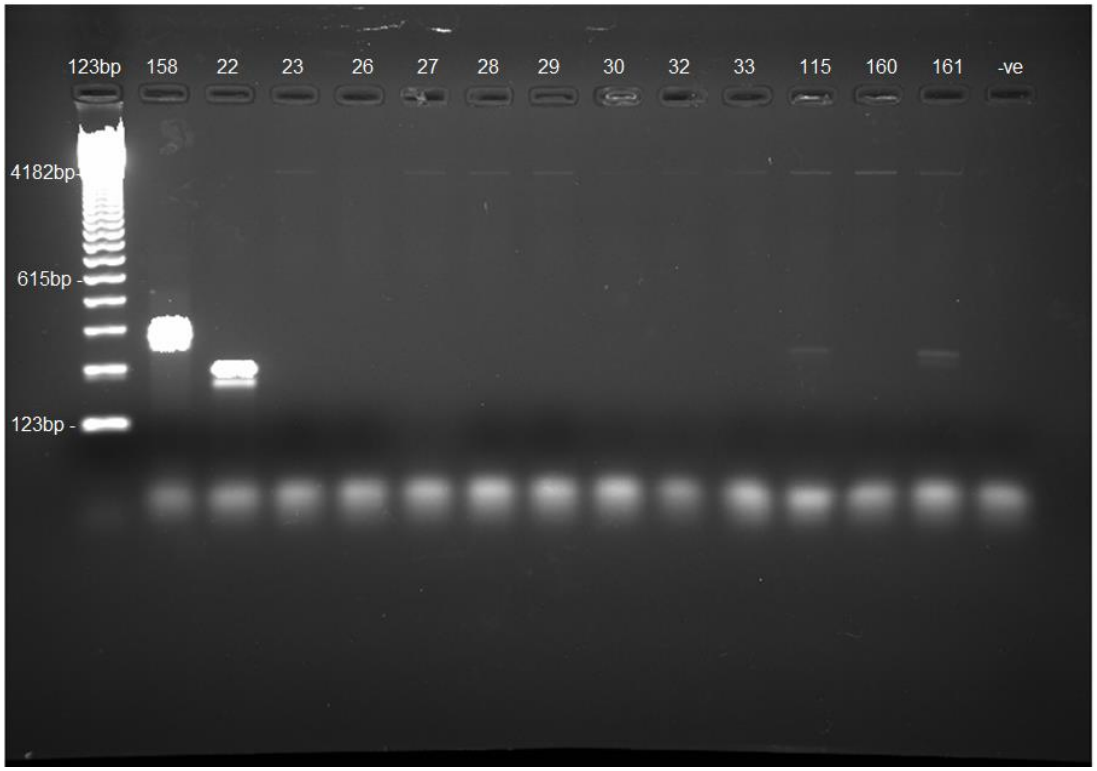
A 100bp and 123bp DNA marker ladder is shown in lanes 1 and 20, respectively. *S. pneumoniae* serotype 33F ID number is listed above each lane. A negative control was included in lane 19 (-ve). A white box surrounds the large >2000bp fragment that was detected in eight of the twelve *S. pneumoniae* isolates.



Since BOX-06 failed to amplify in 75% of Queensland serotype 7F isolates, we verified whether the primers would anneal to the NCBI published genomes. BOX-06 primers appear to anneal to a serotype 7F contig (CDC 1087-00 contig ABFT01000005.1) however the fragment size exceeds 1200bp which would not be detected with the AB3130 internal size ladder LIZ1200. Gel electrophoresis did not reveal any bands, corresponding to BOX-06 in serotype 7F except for three isolates, two of which were very faint bands (Figure 14).

Figure 14: Non-amplification of BOX-06 fragments in *S. pneumoniae* serotype 7F.

A 123bp DNA marker ladder is shown in the first lane, followed by the *S. pneumoniae* isolates with ID number above each lane. Isolate ID 158 was a positive control *S. pneumoniae* serotype 6A with BOX-06 amplification. All isolates in other lanes are serotype 7F, with only isolate ID 22, 115 and 161 with a positive amplification of BOX-06. A negative control was included in the last lane (-ve).



4.3.2 Examination of potential VNTR function in NCBI pneumococcal genomes

In silico examination of the *S. pneumoniae* R6 and TIGR4 strains for positions of VNTR loci used in the various MLVA methods was also performed. Location of the VNTR loci and their proximities to other transcripts was recorded (Table 15).

Table 15: Location of VNTRs in proximity to various proteins and regulators in R6 and TIGR4 *S. pneumoniae* strains.

Closest transcript to the VNTR is at the top of each list ("Upstream of…" and "Downstream of…"). *DNA binding response regulator and sensor histidine kinase work together in a two component regulatory system.

VNTR Locus	*Upstream of…*	*Downstream of…*
BOX-01	NP_359231.1 Catabolite control protein	NP_359229.2 Anthranilate synthase component I
		NP_359228.1 Anthranilate synthase component II
		NP_359227.1 Anthranilate phosphoribosyltransferase
BOX-02	NP_358159.1 β-galactosidase	NP_358158.1 PTS System transporter subunit IIC
		NP_358157.1 Hypothetical protein
		NP_358156.1 PTS system transporter subunit IIA
BOX-03	NP_359196.1 Aquaporin	NP_359195.1 7-cyano-7-deazaguanine reductase
BOX-04	NP_358291.1 hypothetical protein	NP_358290.2 methionyl-tRNA synthetase
	NP_358292.1 hypothetical protein	
BOX-06	NP_359522.1 histidyl-tRNA synthetase	NP_359521.1 hypothetical protein
BOX-10	NP_358172.1 DNA Binding Response Regulator*	NP_359520.1 transcriptional regulator
	NP_358173.1 Sensor Histidine kinase*	
	NP_358174.1 hypothetical protein	
	NP_358175.1 zinc metalloprotease (ZmpB)	
	NP_358176.1 Para-aminobenzoate synthetase	
	NP_358177.1 Pneumococcal surface protein	
	NP_358172.1 DNA Binding Response Regulator	
BOX-11	NP_358009.1 formate acetyltransferase	NP_358008.1 DNA polymerase IV
BOX-12	NP_358962.1 transcription elongation factor GreA	NP_358961.2 hypothetical protein
BOX-13	NP_357687.1 phosphoglycolate phosphatase	NP_357686.1 capsule polysaccharide biosynthesis capD
	NP_357688.1 L-serine dehydratase subunit alpha	
Spneu17	NP_359405.1 Catabolite control protein A	NP_359404.1 L-asparaginase
Spneu19 (overlaps)	NP_359536.1 Choline binding protein PcpA	
Spneu27	NP_357850.1 ribosomal small subunit pseudouridine	NP_357849.1 hypothetical protein
		NP_357848.1 aminopeptidase
Spneu31	NP_358151.1 ABC transporter ATP-binding protein	NP_358150.1 SOS ribosomal protein L1
Spneu39	NP_359526.1 dihydroxy-acid dehydratase	NP_359525.1 hypothetical protein

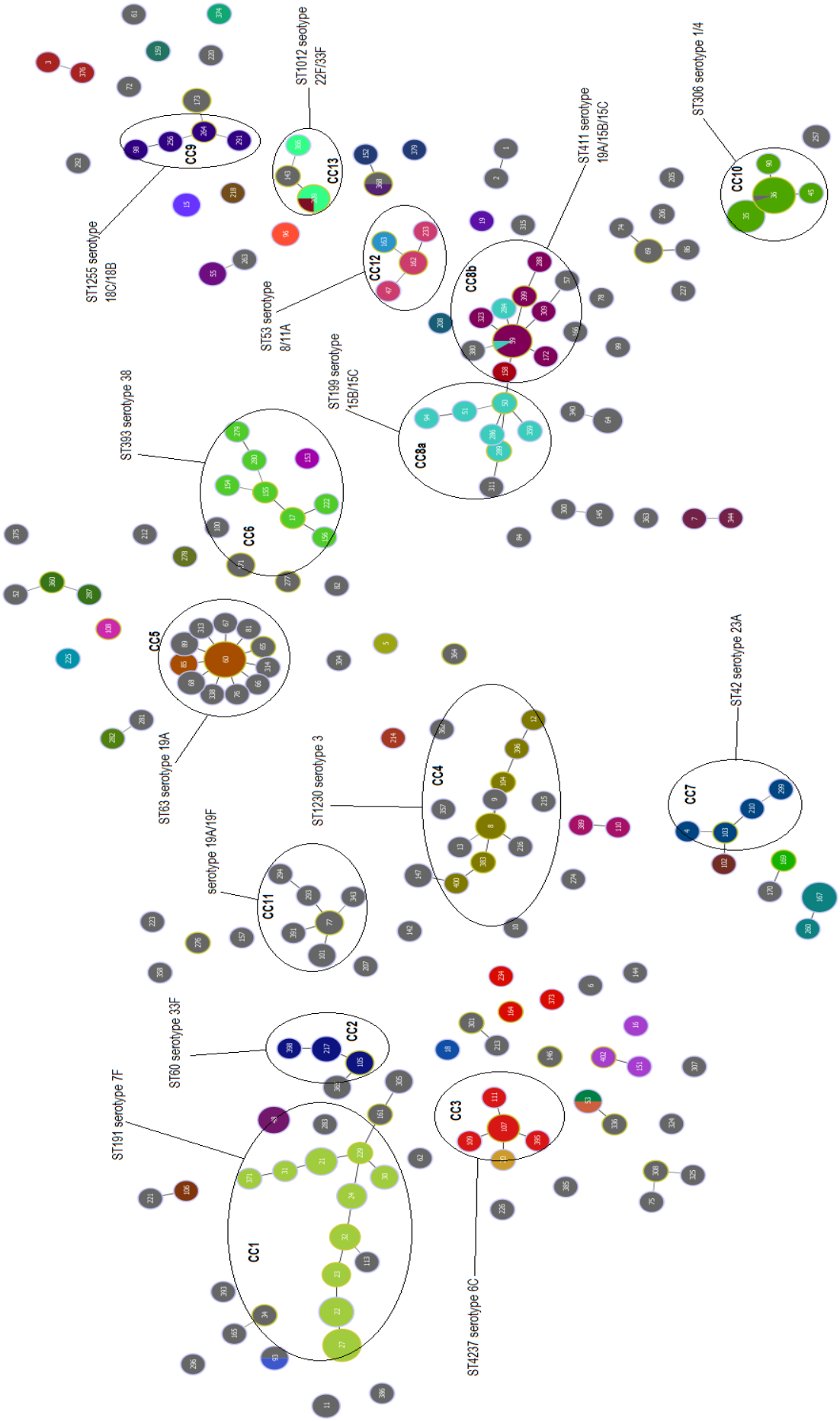
4.3.3 Comparison of MLVA and MLST eBurst

Comparison of MLST and MLVA4 eBurst analysis clearly showed two different population structures (Figure 15). The pneumococcal population structure is displayed according to the MLVA4 genotypes, and it was observed that there are 32 clonal complexes (CC), with the larger eleven complexes labelled in the figure. The MLST results are overlayed on the MLVA4 results; therefore each colour circle represents a different MLST type. It can be seen that several MLVA4 complexes would appear as a single MLST type, for example CC1 which predominantly contains serotype 7F isolates has as many as ten different MLVA4 types however only one MLST type as depicted by the single green colour of the circles.

PLEASE SEE FOLLOWING PAGE FOR FIGURE 15.

Figure 15: Comparison of MLST genotyping to MLVA4 genotyping (see opposite page).

Clonal complexes (CC) are circled and contain single locus variants (SLV) and double locus variants (DLV). MLVA4 genotype is presented inside each circle. Size of each circle represents the number of isolates detected in this study. Grey circles indicate an MLVA4 type but no MLST type. Coloured circles indicate that an MLST type was obtained, and isolates of the same colour have the same MLST type. Circles with split colours indicate that two or more MLST types were obtained for a single MLVA4 type.


ST1012 serotype 22F/33F
ST306 serotype 1/4
ST411 serotype 19A/15B/15C
ST1255 serotype 18C/18B
ST53 serotype 8/11A
ST199 serotype 15B/15C
ST393 serotype 38
ST63 serotype 19A
ST1230 serotype 3
ST142 serotype 23A
serotype 19A/19F
ST60 serotype 33F
ST191 serotype 7F
ST4237 serotype 6C
CC1
CC2
CC3
CC4
CC5
CC6
CC7
CC8a
CC8b
CC9
CC10
CC11
CC12
CC13

Using MLVA4 to type strains of *S. pneumoniae* for localised Queensland studies showed that most of the 32 CCs predominantly contained a common serotype (e.g. CC1 contains serotype 7F, CC4 contains serotype 3). However closer examination of the data revealed that some complexes contained more than one serotype. These complexes included CC11 (serotypes 19A and 19F), CC8 (serotypes 15B, 15C and 19A), CC9 (serotypes 18B and 18C), CC10 (serotypes 1 and 4), CC3 (serotypes 6A and 6C), CC12 (serotypes 8 and 11A), CC13 (serotypes 33F and 22F). Use of MLST identified most of these CCs but many appear as a singleton, meaning only one sequence type was observed.

MLST results were also overlayed with the MLVA1 and MLVA2 results. Both MLVA methods were found to be more discriminatory than MLST, as evident by the diversification of MLST types into many MLVA types (e.g. CC10 contains at least five different MLVA1 types with the same MLST type) (Figure 16 and Figure 17). Interestingly, MLVA1 grouped serotypes 1, 3 and 7F into the same complex (CC10) according to the genotype profile (Figure 16).

Similarly, MLVA2 grouped most of the CCs together as isolates appear to be single locus variants (SLV) or double locus variants (DLV) of another isolate (Figure 17).

PLEASE SEE FOLLOWING PAGE FOR FIGURE 16.

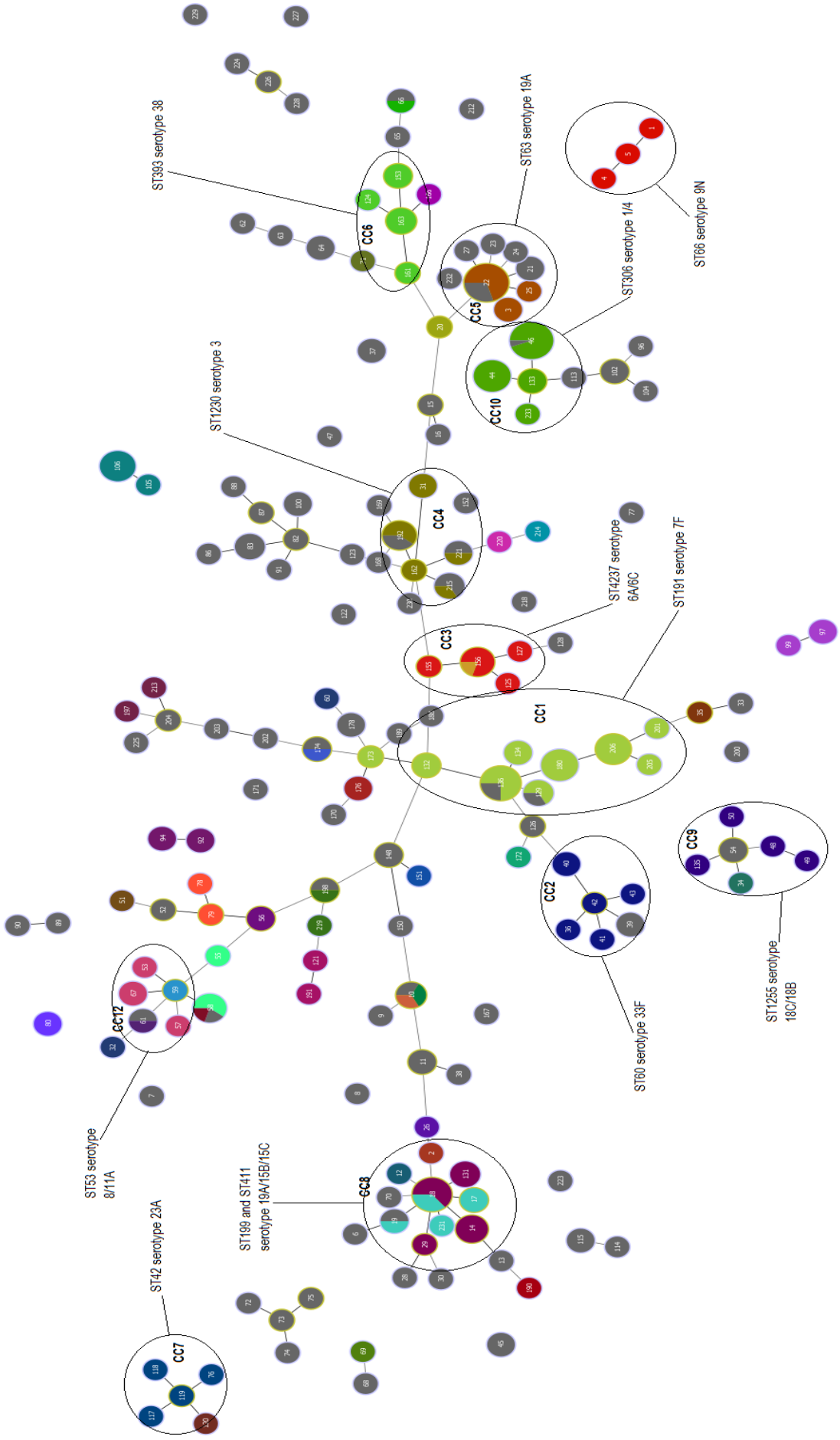
Figure 16: Comparison of MLST genotyping to MLVA1 genotyping (see opposite page).

Clonal complexes (CC) are circled and contain single locus variants (SLV) and double locus variants (DLV). MLVA4 genotype is presented inside each circle. Size of each circle represents the number of isolates detected in this study. Grey circles indicate an MLVA4 type but no MLST type. Coloured circles indicate that an MLST type was obtained, and isolates of the same colour have the same MLST type. Circles with split colours indicate that two or more MLST types were obtained for a single MLVA4 type.


serotype 22F/33F
CC13
ST191
serotype 7F
CC10b
ST1230 serotype 3
St393 serotyp 38
CC6
CC10a
ST4237 serotype
6A/6C
CC3
ST63 serotype 19A
ST306 serotype 1/4
ST411 and ST199
serotype
19A/15B/15C
CC8
ST42 serotype 23A
CC7
ST1255 serotype
18B/18C
ST66 serotype 9N

Figure 17: Comparison of MLST genotyping to MLVA2 genotyping (see opposite page).

Clonal complexes (CC) are circled and contain single locus variants (SLV) and double locus variants (DLV). MLVA4 genotype is presented inside each circle. Size of each circle represents the number of isolates detected in this study. Grey circles indicate an MLVA4 type but no MLST type. Coloured circles indicate that an MLST type was obtained, and isolates of the same colour have the same MLST type. Circles with split colours indicate that two or more MLST types were obtained for a single MLVA4 type.

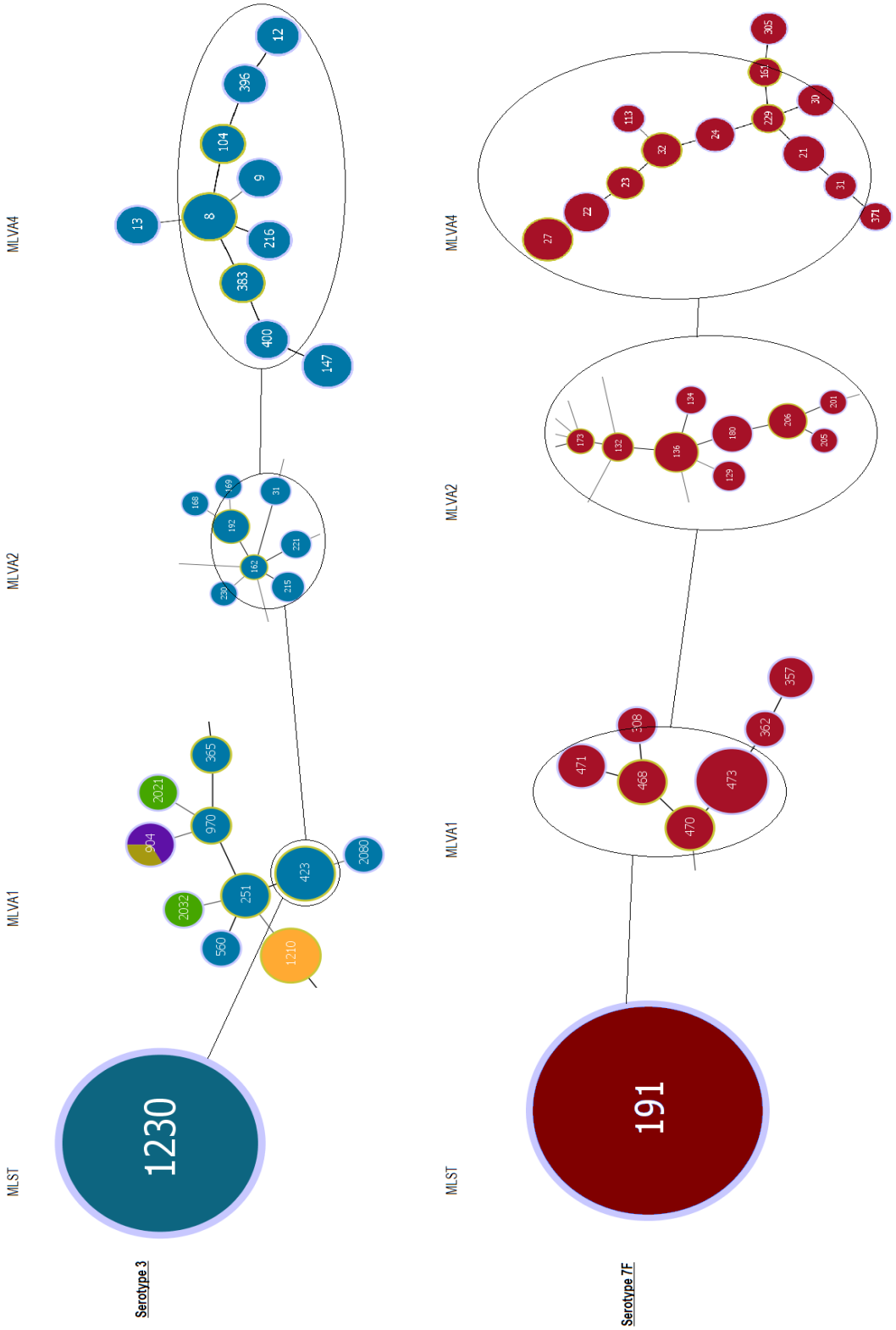

ST393 serotype 38
ST63 serotype 19A
ST66 serotype 9N
ST306 serotype 1/4
ST1230 serotype 3
ST4237 serotype 6A/6C
ST191 serotype 7F
ST1255 serotype 18C/18B
ST60 serotype 33F
ST53 serotype 8/11A
ST199 and ST411 serotype 19A/15B/15C
ST42 serotype 23A
CC1
CC2
CC3
CC4
CC5
CC6
CC7
CC8
CC9
CC10
CC12

However for some serotypes, MLVA4 provides higher discrimination than MLST and other MLVA methods, evident by the high diversification of serotype 7F and serotype 3 (Figure 18). When genotyping *S. pneumoniae* strains with MLST, only a single sequence type is observed for most serotype 7F (ST191) and serotype 3 (ST1230). However these isolates diverge into a number of different MLVA genotypes (all genetically related). MLVA4 identifies at least eight different genotypes for serotype 7F isolates, and thirteen different genotypes for serotype 3. Particular loci that can discriminate serotype 3 further using MLVA4 is Spneu17, and BOX-10 further discriminates serotype 7F.

PLEASE SEE FOLLOWING PAGE FOR FIGURE 18

Figure 18: Comparing the population structure of *S. pneumoniae* serotype 3 and serotype 7F when using MLST, MLVA1, MLVA2 and MLVA4 (see opposite page).

The MLST genotype for serotype 3 (blue) and serotype 7F (maroon) is shown to diversify (i.e. break into distinctive genotypes) when using MLVA1, MLVA2 and MLVA4. The respective genotype is presented inside each circle (e.g. MLST identifies ST1230, whereas MLVA1 identifies MT423). Other coloured circles depict other serotypes that were genetically related to serotype 3 according to MLVA1.


MLST
MLVA1
MLVA2
MLVA4
1230
365
2021
904
970
2032
251
560
1210
423
2080
168
169
192
31
221
162
215
230
12
396
104
9
13
8
216
383
400
147
Serotype 3
MLST
MLVA1
MLVA2
MLVA4
191
357
362
308
471
468
473
470
305
161
30
229
113
24
32
21
23
31
22
371
27
134
173
132
136
180
206
201
205
129
Serotype 7F

4.3.4 Capsule switching

MLST and MLVA4 used in combination with serotyping can identify potential capsule switching. MLST identifies seven potential capsule switches including a switch from serotype 22F to serotype 33F (ST1012), a switch from serotype 18B to serotype 18C (ST1255), a switch from serotype 15B to serotype 15C (ST199), a switch from serotypes 1 to serotype 4 (ST306), a switch from serotype 19A to serotype 15B and serotype 15C (ST411), a switch from serotype 6A to serotype 6C (ST4237) and a switch from serotypes 8 to serotype 11A (ST53) (Table 16). On the other hand, MLVA4 only detects two capsule switches including serotypes 15C to 19A (MT59; ST411), and serotypes 1 to 4 (MT36; ST306). The Queensland *S. pneumoniae* isolates are listed in the supplementary data which contains MLVA1, MLVA4, MLST and serotype for comparison.

Table 16: Potential capsule switches detected by MLST, MLVA1 and MLVA4 methods. The potential capsule switches (boxed) have been consistently detected with all three genotyping methods.

MLST (ST)	**Serotype (year observed); number of isolates**
1012	22F (2007-2008); n=3 33F (2012);n=1
1255	18B (2007);n=1 18C (2008,2010);n=3
199	15B (2007,2010,2011);n=4 15C (2007,2010,2012);n=4
306	1(2007-2012);n=23 4(2012);n=1
411	19A (2007-2012);n=12 15B (2009);n=1 15C (2007);n=2
4237	6A (2008);n=1 6C (2010-2012);n=6
53	8 (2008-2009);n=3 11A (2011);n=1
MLVA1	**Serotype (year observed); number of isolates**
1190	19A (2007-2012);n=3 19F (2007,2010);n=2
1374	6A (2008);n=1 6C (2010-2012);n=5
18	8 (2008-2009);n=3 11A (2011);n=1
2002	18B (2007);n=1 18C (2008-2010);n=3
261	1 (2007-2012);n=15 4 (2012);n=1
58	15B (2011);n=1 15C (2007, 2012);n=3

MLST (ST)	Serotype (year observed); number of isolates
67	15C (2007);n=1 19A (2007-2012);n=12
904	15B (2012);n=1 15C (2011);n=2
MLVA4	**Serotype (year observed); number of isolates**
59	15C (2007) ; n= 1 19A (2007); n = 10
36	1 (2007-2012); n=16 4 (2012); n= 1
107	6A (2008); n= 1 6C (2010-2012); n= 3

Finally, Van Cuyck *et al.* (2012) suggested that MLST sequence types could be inferred from the MLVA type, although they commented that serotype 19F could not be inferred. MLVA4 and the Queensland MLST types were examined to determine whether some ST could be inferred (Table 17). Due to the fact that several sequence types cannot be inferred from a single MLVA4 type, it is not advisable to adopt this method of inferring one genotype from another.

Table 17: Identification of multiple MLST types to a single MLVA4 type when genotyping *S. pneumoniae.*

Serotype	Sequence type (MLST)	MLVA4 type
15C	1262; 5216	53
15B/15C/19A	199; 11	59
11A	557; 9668	93
22F/33F	1012; 5599	209

4.4 Discussion

MLST is considered the 'gold standard' genotyping method for studying invasive *S. pneumoniae*. However MLVA has also emerged as an alternative genotyping technique as it has higher discriminatory power, is quick and inexpensive than MLST (Koeck *et al.,* 2005; Elberse *et al.,* 2011a). Results presented in this study support this verdict, although MLVA would be more suitable for short-term and localised/state-wide epidemiology studies rather than long term and global studies because of the higher rate of diversification of VNTR loci compared to housekeeping genes used for MLST. Currently there are a variety of MLVA protocols for typing *S. pneumoniae*, some of which fail to amplify a number of loci in different serotypes; hence this study describes the development of a modified MLVA4 method in an attempt to resolve this

issue as well as to maintain a highly discriminatory, inexpensive and quick genotyping technique.

Variations in interpreting the population structure of *S. pneumoniae* have been observed when using different genotyping protocols (MLVA1, MLVA2, MLVA4 and MLST methods). The choice of the ten VNTRs for our MLVA4 method was based on a Hunter-Gaston Diversity of ≥0.8, as well as an anchor locus with low discrimination to determine long-term changes (BOX-02), and an extra locus with high discriminatory power for specific serotypes (Spneu19). Of the eight highly discriminatory (Hunter-Gaston Diversity ≥0.8) VNTR loci used in previously published MLVA papers (Elberse *et al.,* 2011a; Koeck *et al.,* 2005; Rakov *et al.,* 2011; Van Cuyck *et al.,* 2012), seven VNTRs are included in the modified MLVA4 method. MLVA4 maintains high congruence with the other MLVA methods and MLST.

MLST proved to be less discriminatory (S=0.936) than either of the MLVA1 (S=0.963) and MLVA4 methods (S=0.978). However MLST remains as an ideal method for long-term and global epidemiology studies as the housekeeping genes change slowly over a longer period of time (Bentley *et al.,* 2006). A less discriminatory protocol is only problematic if it does not detect emerging genotypes or outbreaks therefore inaccurately representing the pneumococcal population structure, justifying why MLVA4 may be more suitable.

MLVA4 method maintains a high discriminatory power whilst minimising the number of non-amplifiable '99', and is significantly more accurate in representing the Queensland pneumococcal population structure compared to MLST when observing short-term changes or localised outbreaks. Complexes of isolates can be observed in more detail and more information can be gleaned when combined with isolate information such as antibiotic resistance or disease type (not used in this study). Admittedly, additional markers have increased resolution (ten loci instead of seven or eight) however we have minimised laboratory work to three multiplex reactions. The main purpose of modifying MLVA was to eliminate or minimise the number of non-amplified loci as detected in the Elberse *et al.* (2011a) method, and this has been achieved in this study.

Results of the present study confirm the problem of particular loci failing to amplify, resulting in incomplete genotypes. Non-amplification could also be due to primer-dimer forming in the mastermix, therefore alternative combination or pooling of primers could be performed. Alternatively a mutation in the primer binding site or a missing locus could explain non-amplification. It was noticed that singleplex PCR resolved >70% of certain loci, indicating that the genes can be amplified but a multiplex

may be inhibiting amplification in MLVA1. A number of primers were also re-designed in this study (including BOX-10, BOX-12, BOX-13 and Spneu19) in an attempt to successfully amplify the previous failing.

Elberse *et al.* (2011b) reported that 24% of their *S. pneumoniae* isolates still contained one or more non-amplified BOX loci even after repeated rounds of PCR on these isolates. Results from the present study provided a similar figure of 20.8% in Queensland isolates. It also found that some loci failed to amplify in specific serotypes, such as BOX-06 failing to amplify in 75% of serotype 7F, indicating that primers should be redesigned, or VNTR fragments were absent. No serotype 7F genome is available in the NCBI database. Elberse *et al.* (2011b) similarly have detected a large number of non-amplified BOX-06 genes associated with serotype 7F (89% isolates). BOX-06 was not used in MLVA4 due to a high percentage of '99' associated with serotype 7F and a low Hunter-Gaston diversity. Since serotype 7F was the second most common serotype (9%) found in Queensland, using higher discriminatory loci was favoured. Croucher *et al.* (2011) identified that "the greater the average number of *boxB* repeats in a BOX element, the more likely that element is to vary by losing or acquiring these modules". This may explain why it has been difficult to amplify certain loci because they are less stable and have the capacity to transpose more easily. In the present study it was also noticed that MLVA2 had similar numbers of non-amplification, with 24.3% of Queensland isolates with incomplete MLVA genotypes. MLVA4 was able to decrease this percentage to 12.4% of Queensland isolates, and the non-amplifications were thought to be serotype-specific. If the absence of these BOX loci is real for these serotypes, this could be a discriminatory feature.

Similarly, Spneu19 loci could not be detected in several serotype 3 genomes in the NCBI database, explaining why experimentally Spneu19 is assigned '99' for several serotype 3 isolates. The failure to amplify Spneu19 in serotype 3 has also been observed by Koeck *et al.* (2005), suggesting that serotype 3 lack *pcpA* which codes for a non-essential surface protein involved in cell adhesion (Koeck *et al.,* 2005; Sanchez-Beato *et al.,* 1998). On the other hand, large BOX-13 fragments (>2000bp) have been identified in serotype 33F isolates, accounting for the '99' results since the AB3130 internal size ladder only reaches 1200bp. Large fragments could be explained by the placement of an insertion sequence (IS) element, making the BOX element appear to be larger than 2kb. The presence of IS elements has been described in other MLVA studies (Hoang *et al.,* 2007). Further experiments could investigate the use of other types of VNTRs for example RUP (Repeat Units of Pneumococcus) and SPRITE elements. However little is known about these elements in *S. pneumoniae.* RUP elements have

108bp repeat units and are located in *recA* – *dinF* intergenic regions (Claverys & Martin, 1998; Oggioni & Claverys, 1999). There are 32 copies of RUP found in the NCBI *S. pneumoniae* R6 strain genome and are thought to be insertion sequence (IS) derivatives that could be mobile (Oggioni & Claverys, 1999).

Previous studies have investigated the question of whether or not BOX/VNTR loci are stable *in vitro* and *in vivo* (Elberse *et al.,* 2011a; Martin *et al.,* 1992). The stability of repetitive sequences and PCR patterns was observed in six species of bacteria after 72hours and 15 sub-cultures (although this was investigated in other bacterial species, not *S. pneumoniae*) (Kang & Dunne, 2003). Importantly, Elberse *et al.,* (2011c) demonstrated BOX stability in *S. pneumoniae* after sub-culture of the organism for 29 days. Sixteen VNTRs were also shown to be stable in *Mycobacterium intracellulare* strains when 30 sub-cultures over 180 days demonstrated *in vitro* stability and that 14 patients' isolates recovered from 2 – 1551 with identical MLVA profiles at each interval demonstrated *in vivo* stability (Ichikawa *et al.,* 2010). MLVA performed for *Acinetobacter baumannii* provided stable VNTR profiles after two sub-cultures (Pourcel *et al.,* 2011). McAuliffe *et al.* (2007) also demonstrated MLVA stability in *Mycoplasma mycoides* when three strains were sub-cultured 50 times. All these studies have demonstrated that within a short experimental time frame, at least, VNTR loci are stable and MLVA profiles should accurately reflect the organism's genotype.

Out of interest, examination of potential VNTR function was investigated by inspecting the NCBI database of *S. pneumoniae* strains R6 and TIGR4 as examples. It was interesting to note the close proximity the majority of VNTR loci have to other transcribable elements. BOX-01 was located upstream of a catabolite control protein, which functions in colonization (Iyer *et al.,* 2005). BOX-02 was located upstream of β-galactosidase, which catalyses the hydrolysis of β-galactosidases into monosaccharides. Upstream of an aquaporin, which forms pores in a membrane, was BOX-03. BOX-10 was located upstream of a number of transcriptional regions. For instance, it was located near a two-component regulatory system (DNA binding response regulator/sensor histidine kinase) which allows the organism to sense and respond to changes in different environmental conditions. External stimuli activate histidine kinase which transfers a phosphoryl group to a response regulator (Stock *et al.,* 2000). Another possible function of BOX-10 may be to effect zinc metalloprotease *ZmpB*, which requires zinc metal with the enzyme protease to function in proteolysis (the breakdown of proteins into smaller peptides/amino acids). Zinc metalloprotease strongly contributes to virulence (Chiavolini *et al.,* 2003). Also, BOX-11 was found upstream of formate acetyl transferase which catalyses acetyl-CoA and formate to CoA

and pyruvate. It would be interesting to further investigate whether or not the VNTR elements affect these particular downstream elements.

MLVA4 combined with serotyping, increases our understanding of the Queensland pneumococcal epidemiology. Serotyping alone identifies 35 serotypes but prohibits our understanding of genetic relationships and capsule switching. MLST identifies eleven clonal complexes (n=202) circulating throughout Queensland, whereas MLVA4 identifies eighteen clonal complexes (n=202) and also enables closer examination of eleven MLST singletons, one of which is serotype 7F which is commonly associated with IPD in Australian Aborigines. Further discussion of the Queensland pneumococcal population is discussed in Chapter 5. Unfortunately the Queensland MLST results cannot contribute to a national Australian study/database because none currently exists. Few Australian isolates have been submitted to the international MLST database, and we have submitted the MLST results from this study to the MLST database. MLVA4 reveals diverse genetic relationships within the Queensland population, including between serotypes 18B and 18C (CC7), serotypes 19A, 15C and 15B (CC6), serotype 8 and 11A (CC21), serotype 7F (CC1) and serotype 3 (CC3) which MLST does not reveal. Overlaying patient information with diverse genotypes may reveal further information about outbreaks, patient susceptibility, vaccine immunisation and tracking of disease.

Genetically related isolates of *S. pneumoniae* may engage in capsule switching more frequently than other bacteria. Potential capsule switches have already been observed between serotypes 19A and 15C in CC6 (MT59; ST411), and serotypes 1 and 4 (MT36; ST306) in CC8 in our Queensland population using MLVA4 (Table 16). MLST also identifies these capsule switches, as well as many others which may indicate false capsule switching since MLST is less discriminatory in discerning the true genetic background of *S. pneumoniae* isolates. Alternatively, MLVA4 is also noted as being too discriminatory and even though true capsule switching occurs, MLVA4 identifies two distinct genotypes therefore the assumption is that no capsule switching has occurred. Further investigation is required to determine whether MLVA4 could fail to detect capsule switches or whether MLST detects false capsule switches. Since MLVA4 is highly discriminatory for short-term studies, it may enable detection of capsule switching earlier than MLST would.

Accurately and quickly detecting relationships between *S. pneumoniae* serotypes may have an impact on the selection of serotypes for future vaccine strategies. Ideally vaccine strategies should benefit the whole of a nation, however differences in state pneumococcal populations have prompted localised strategies, for

example the 10vPCV was only introduced into Northern Territory in October 2009 since the additional serotypes were not common elsewhere in Australia. If a vaccine targets only one serotype in a CC and not the other, this may provide pressure for capsule switching or serotype replacement. The 13vPCV targets serotype 19A, therefore there might be a selective pressure for the serotype 19A to switch capsules to 15C or 15B (CC6), which are known to be genetically related. Serotype 15B and 15C are not included in any current childhood vaccines, or the in-trial 15vPCV which targets the extra serotypes 22F and 33F, therefore future surveillance is recommended. It is unknown how many serotypes in a vaccine are acceptable for long-term decrease of IPD.

Furthermore, genotyping methods have revealed that certain pneumococci of a particular genotype are more commonly associated with invasive disease. For instance, serotype 1 is a prevalent pneumococcal type in Australia - this has been widely established via serotyping (e.g. Quellung reaction). However with the use of MLST genotyping, we have gained further understanding that particular serotype 1 strains are more commonly associated with virulence and disease than others, for instance ST306 (serotype 1) is very prominent (Staples *et al.,* 2015). It is unknown why different genotypes of serotype 1 are associated with virulence more than others; however research has shown that genetic pneumococcal clones have displayed heterogeneous behaviour in invasiveness (Sá-Leão *et al.,* 2011). If we use MLVA4 genotyping, because it is more discriminatory than MLST, the pneumococcal isolates are further distinguished by their MLVA genotype; hence serotype ST306 is also associated with several different MLVA types, of which MT35 and MT36 appears to be more common with invasiveness. This demonstrates the importance of molecular surveillance, particularly following the introduction of vaccines, as not only serotypes can be associated with invasiveness but the genotypes (either MLST or MLVA) can also be associated with virulence. This could then inform not only which serotype but even which specific strain/genotype could be targeted.

In addition, the ability to examine CC with higher discrimination using MLVA4 can provide insight into which genes are changing, for example serotype 7F (MLVA CC1 or ST191) largely diversifies due to BOX-10 and serotype 3 diversifies due to Spneu17. Investigation into specific VNTR loci may indicate whether there are additional advantages that these serotypes may have based on the VNTR proximity to virulence elements. This could offer these isolates a selective advantage over others (Table 15). It is unknown what the specific functions of these genes are, however it is known that VNTRs and BOX elements play a role in bacterial competence and virulence (Knutsen *et*

al., 2006; Martin *et al.,* 1992). VNTR loci with high diversity (e.g. Spneu17) would allow increased discrimination within localised or short-term studies, whereas VNTR loci with low diversity (e.g. BOX-02 and BOX-11) would allow identification of long term changes.

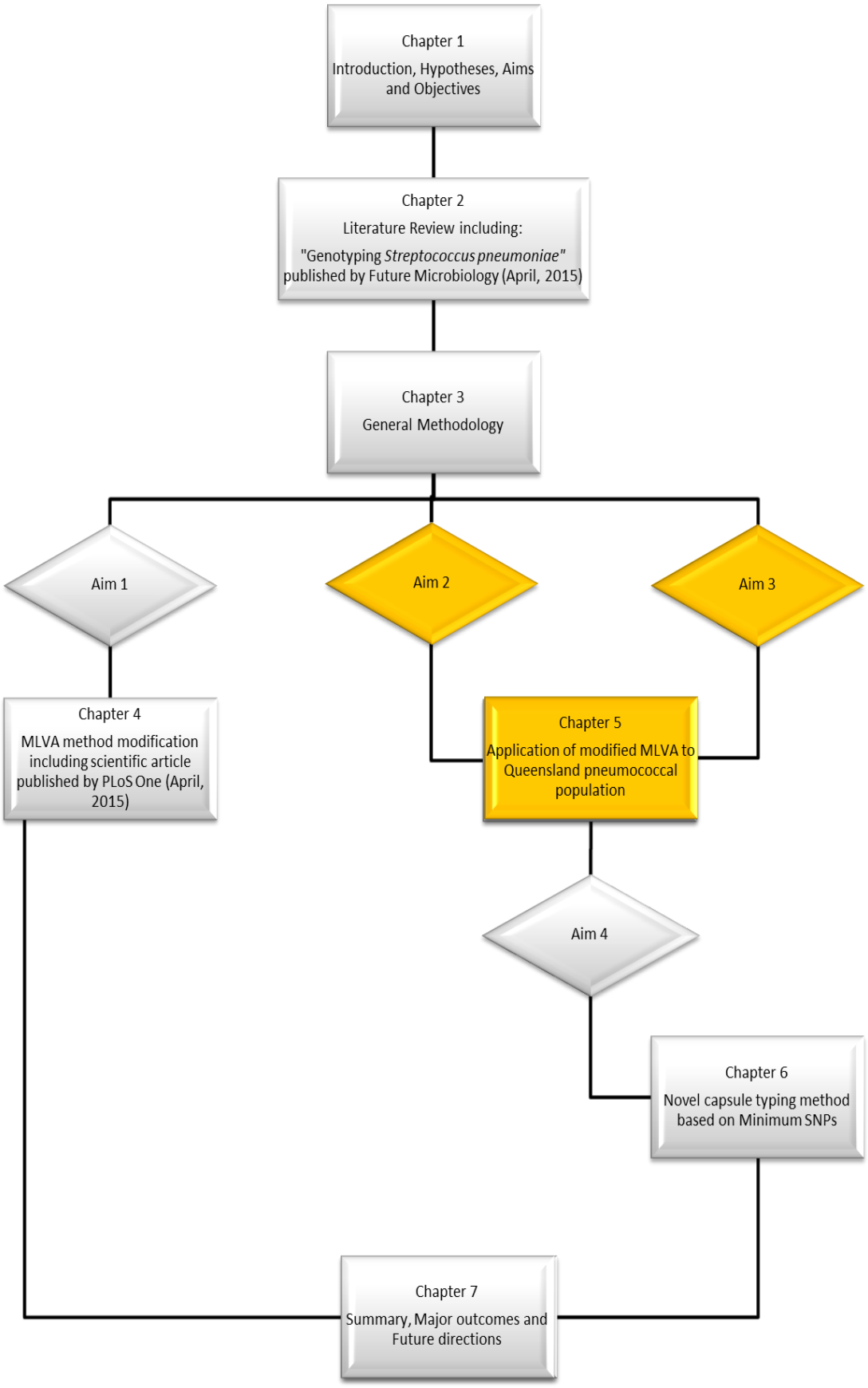
MLVA4 also identified founding genotypes in Queensland whereas MLST did not, however this is only a probable founder based on an eBurst model, and the more data available the more confident about the founder assignment, therefore it is advised to use the whole database if possible. MLST may require a larger population sample or national/international sample before founding genotypes are discerned, therefore the international database has been used for comparison. Unfortunately MLVA4 has no larger database to compare to, although MLVA1 has an international database available with predominantly isolates from the Netherlands. Identification of founding genotypes provides a preliminary understanding of the originating/parent genotype and hence how CCs may have evolved over time. The MLST eBurst limits our understanding of these changes occurring over a short period of time within the state of Queensland. It would be possible to overlay geographic information with genotypes in order to track the spread of specific pneumococcal strains.

In conclusion, we have developed a MLVA4 method for genotyping invasive strains of *S. pneumoniae.* The main advantage of this new method over other MLVA methods for typing *S. pneumoniae* is the ability to achieve complete MLVA profiles for serotypes whilst also maintaining a highly discriminatory and quick genotyping method. Loci of invasive serotypes of *S. pneumoniae* that failed to amplify were found to be serotype specific, which may indicate that these BOX elements in these serotypes are unstable and have the capacity to transpose. Further research is required to understand the VNTR genetics of these serotypes as VNTRs and BOX loci are thought to play a role in virulence. MLVA4 is also more suitable for genotyping *S. pneumoniae* in Queensland than MLST as a more diverse population can be visualised and allows accurate tracking of strains across the state. MLST may be more suitable for a national study, rather than by state, however as MLVA4 is more discriminatory it is possible to utilise this for national studies also as Australia-wide outbreaks have been observed e.g. pneumococcal ST1. MLVA4 has been applied to analyse the pneumococcal population in Queensland from 2007 to 2012. This study, following the introduction of 13vPCV in 2011, establishes a baseline and it is expected that future monitoring will comprehensively and accurately depict the changes in the pneumococcal population in Queensland. Currently, no MLVA database exists for *S. pneumoniae* however there is an opportunity for this to be established. The future perspective of MLVA is that it will

emerge as an inexpensive and quick genotyping method for localised and national studies that can be used in conjunction with the currently traditional and slower serotyping and MLST methods for characterising *S. pneumoniae*.

A low numbers of available sequences, transposable elements make it difficult to design suitable primers, and selection pressures can lead to genetic variation of two identical clones (Lindstedt, 2005). MLVA4 has not been tested directly on clinical samples (without culture), but has the potential to obtain similar results as MLST (Elberse *et al.,* 2011a). It is also limited by size homoplasy – the "independent evolution of the same repeat array sizes in different strains" (Rakov *et al.,* 2011). MLVA profiles would not be able to differentiate between independently evolved strains with the same profile. However, the choice of the 45bp *boxB* tandem repeat sequence means that it is less prone to size homoplasy (Rakov *et al.,* 2011).

CHAPTER 5: CHANGING POPULATION STRUCTURE OF INVASIVE *STREPTOCOCCUS PNEUMONIAE* IN QUEENSLAND CHILDREN


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

5.1 Introduction

The pneumococcal population structure causing IPD in Queensland children could be determined using the modified MLVA4 method developed in Chapter 4. Despite a decline in circulating 7vPCV-serotypes worldwide and associated IPD due to the pneumococcal vaccines, non-7vPCV serotypes started to increase globally due to two phenomena; serotype replacement and capsule switching (Brueggemann *et al.,* 2007; Weinberger *et al.,* 2011). Serotype replacement is the main reason for the shifting pneumococcal population structure. Alarmingly, capsule switching events have led to vaccine escape pneumococcal isolates, where strains with genetic backgrounds providing robustness, virulence, antibiotic resistance, etc. are switching from 7vPCV to non-7vPCV capsules to escape vaccines, and that this is leading to higher rates of non-7vPCV disease/carriage. No vaccine escape isolates of pneumococci have been reported in Australia; however serotype replacement in these bacteria has been reported (Hanna *et al.,* 2010). IPD notification rates in Queensland have increased from 6.0 per 100 000 to 7.6 per 100 000 from 2009 to 2012 (NNDSS, 2015). This is possibly due to the increase of non-7vPCV serotypes in Queensland. Young children (under 2 years) and the elderly (over 65 years) are more susceptible to pneumococcal infections (Scott *et al.,* 1996).

To combat changing pneumococcal populations and increasing invasive diseases, a 13vPCV was developed (Pfizer Australia Pty Ltd) to replace the 7vPCV. This decision and selection of serotypes was based on international pneumococcal serotype surveillance data (Centers for Disease Control Prevention, 2010). In July 2011 the 13vPCV was added to the National Vaccine Immunisation Schedule in Australia, targeting an extra six serotypes not covered by the earlier 7vPCV (Selvey, 2011). Since the introduction of 13vPCV in Australia, there have been no publications examining the effect of this new vaccine on the pneumococcal population structure in Queensland or Australia. A 15vPCV is currently on trial targeting an extra two serotypes (serotype 22F and 33F), and may be a potential future vaccine (Skinner *et al.,* 2011).

The second aim of this thesis was to determine the population structure of invasive *S. pneumoniae* in Queensland children younger than 15 years from January 2007 to December 2012 using the modified MLVA4 method, and also to determine whether or not the genetic diversity of non-vaccine serotypes (non-13vPCV serotypes) is changing in the Queensland pneumococcal population. The population structure will be determined by using the modified MLVA4 method from Chapter 4 and the 'gold standard' MLST method. A number of objectives will be fulfilled in this current chapter.

Firstly, the objective was to determine the differences when interpreting the population structure when using either genotyping method. The second objective was to provide a current study and baseline of the Queensland invasive pneumococcal population in children prior to the introduction of the 13vPCV. Thirdly, research outlined in this Chapter aims to assess the initial and possible future impact of the newly introduced 13vPCV into Queensland by examining the potential changes in the population structure. The final objective was to possibly predict which non-vaccine serotypes may become serotype replacements, indicating that a new vaccine may be required in the future. We hypothesise that the population structure of invasive *S. pneumoniae* isolated from Queensland children are not clonally related from post-7vPCV to post-13vPCV, meaning that the majority of the pneumococcal isolates will not share similar genetic profiles and form genetic complexes as demonstrated by MLVA4 typing and MLST. The rationale for this is that due to the introduction of the 13vPCV, the vaccine is expected to provide pressure on the population structure and we expect to observed differences over time. Finally we hypothesise that serotype replacement and capsule switching has occurred since the freely available 7vPCV in Australia 2005.

5.2 Materials and Methods

5.2.1 Laboratory methods

S. pneumoniae isolates were serotyped (n=397) using Quellung reaction at the Pneumococcal Reference Laboratory, QHFSS. Invasive *S. pneumoniae* isolated from children 15 years or younger from January 2007 to December 2012 were used in this study. The introduction of the 13vPCV occurred July 2011, which was included in the National Immunisation Program.

MLVA1 was performed as previously described in Chapter 3 (Elberse *et al.,* 2011a). Briefly, eight fluorescently labelled BOX MLVA primers were used to amplify invasive *S. pneumoniae* isolates using PCR and Qiagen Mastermix reagent (Qiagen, Australia). An AB3130 sequencer (Applied Biosystems, Australia) was used to size PCR fragment products with a LIZ1200 internal size ladder (Applied Biosystems, Australia).

Our modified MLVA4 method was also applied as described in Chapter 3 and Chapter 4. Briefly ten fluorescently labelled VNTR primers were used to amplify invasive *S. pneumoniae* isolates using PCR, and Qiagen Mastermix reagent (Qiagen, Australia). An AB3130 sequencer (Applied Biosystems, Australia) was used to size PCR fragment products with a LIZ1200 internal size ladder (Applied Biosystems, Australia).

MLST was also performed on selected isolates, namely non-13vPCV serotypes. These isolates were selected because important changes in the pneumococcal population structure may be associated with these isolates. MLST was performed as previously described in Chapter 3 (Enright & Spratt, 1998), using the Corbett Cas 1200 liquid handling robot (Qiagen, Australia) for automated sample and mastermix additions (Jefferies *et al.,* 2003).

5.2.2 Analysis

MLVA results were analysed on PeakScanner V1.0 software (Applied Biosystems, Australia). A MLVA1 type (MT) was assigned to each isolate using the MLVA database (http://www.mlva.net) (Schouls & Van der Heide, 2012). For MLVA4 method, a database was devised and MT types were assigned to each isolates from this database. The population structure for both MLVA methods were analysed by PhyloViz analysis software and displayed as an eBurst diagram.

MLST results were analysed using ChromasPro software (Technelysium Pty Ltd.) for batch alignment analysis. Sequence types (ST) were assigned to each isolate using the MLST database. The MLST population structure was displayed as an eBurst diagram.

Simpson's Index of Diversity (D) was also calculated to compare the discrimination ability of MLVA and MLST methods (accessed from http://alyoung.com/labs/biodiversity_calculator.html).

5.3 Results

5.3.1 Epidemiology of invasive pneumococci

A total of 397 invasive *S. pneumoniae* isolates that had been isolated from children under 15 years old in the years 2007 to 2012 were stored at the pneumococcal reference laboratory, QHFSS. A total of 317 invasive *S. pneumoniae* isolates were genotyped using the MLVA1 and MLVA4 method. MLST was used to only sequence 202 invasive *S. pneumoniae* isolates due to time and cost limitations.

When examining invasive pneumococcal isolates from children 15 years or younger that had been referred to the laboratory, a total of 35 different serotypes were associated with invasive disease. The most common pneumococcal serotypes in this study were serotypes 1 (9%), 3 (5%), 6C (3%), 7F (9%), 15C (3%), 19A (36%), 19F (4%) and 33F (3%) (Table 18). Of these isolates, 9% (serotype 6C, 15C and 33F) were not targeted by the 13vPCV. It was noticed that 68.5% of isolates were associated with children under five years old (Table 18). The total number of pneumococcal isolates has appeared to decline from 37% (147/397) to 31% (125/397) from 2007 to 2012 in children under 5 years, however isolates have increased in older children from 11% (44/397) to 30% (81/397) from 2007 to 2012, although this does not appear to be significant (p=0.1301) (Table 18).

Table 18: The distribution of *S. pneumoniae* serotypes causing invasive disease in Queensland children age <5 years and >5 years in the time periods of 2007-2009 and 2010-2012.

Number of isolates, N; percentage of isolates, (%).

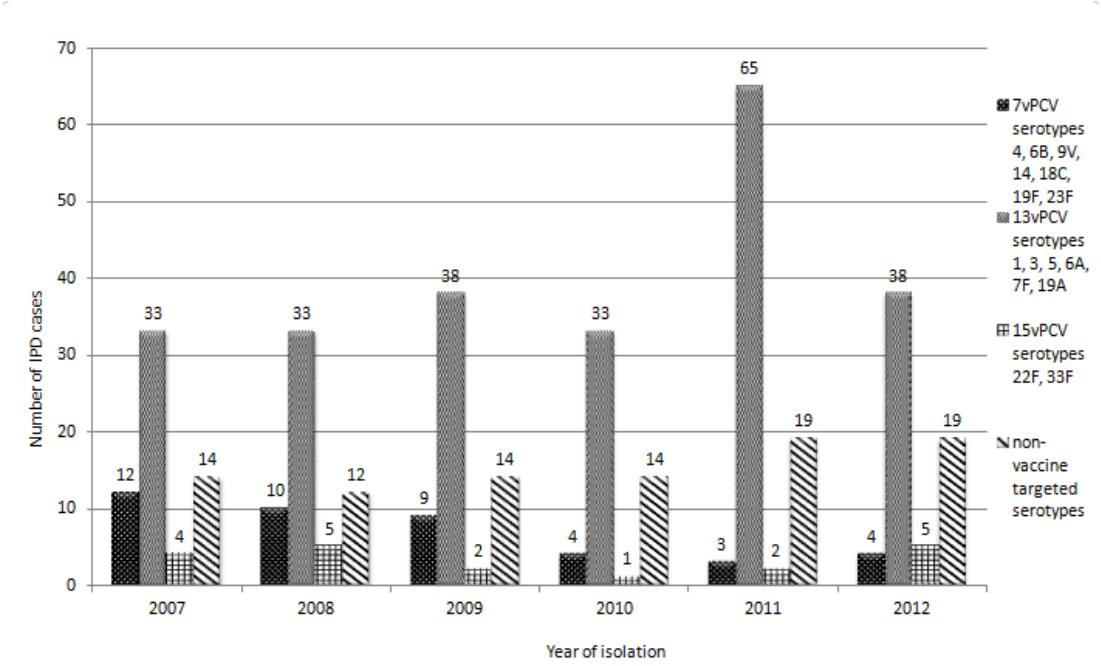
	Pneumococcal Isolation Year				
	2007-2009		2010-2012		TOTAL
	<5 yr	≥5 yr	<5 yr	≥5 yr	
Serotype	N (%)	N (%)	N (%)	N (%)	N (%)
1	1 (0.7)	9 (20)	6 (5)	19 (23)	35 (9)
3	7 (5)	2 (5)	5 (4)	5 (6)	19 (5)
4	1 (0.7)	2 (5)	1 (0.8)	1 (1)	5
6A	3 (2)	1 (2)		1 (1)	5
6B	2 (1)				2
6C		1 (2)	4 (3)	7 (9)	12 (3)
7F	4 (3)	4 (9)	16 (13)	12 (15)	36 (9)
8	4 (3)				4
9N	2 (1)			1 (1)	3
9V	2 (1)			1 (1)	3
10A	1 (0.7)	2 (5)			3
10F	2 (1)				2
11A	5 (4)		1 (0.8)	2 (2)	8
12F		2 (5)	1 (0.8)	1 (1)	4
14	2 (1)	1 (2)		1 (1)	4
15B	2 (1)		3 (2)	1 (1)	6
15C	4 (3)	1 (2)	7 (6)		12 (3)
16F				1 (1)	1
18A	1 (0.7)	1 (2)	1 (0.8)	3 (4)	6
18B	1 (0.7)				1
18C	2 (1)	3 (7)	1 (0.8)		6
19A	68 (48)	5 (11)	57 (46)	15 (19)	145 (36)
19F	8 (6)	1 (2)	5 (4)	1 (1)	15 (4)
22A	1 (0.7)				1
22F	5 (4)		1 (0.8)	1 (1)	7
23A	1 (0.7)	1 (2)	1 (0.8)	2 (2)	5
23B	1 (0.7)		3 (2)	3 (4)	7
23F	3 (2)	4 (9)			7
24F	1 (0.7)				1
33B			1 (0.8)		1
33F	4 (3)	2 (5)	6 (5)		12 (3)
34	1 (0.7)			1 (1)	2
35B	2 (1)		2 (2)	1 (1)	5
35F	1 (0.7)		2 (2)	1 (1)	4
38	5 (4)	2 (5)	1 (0.8)		8
TOTAL	147	44	125	81	397

An increase of 13vPCV-serotypes from 33 isolates (63%) in 2010 to 65 isolates (73%) in 2011 was observed, however this was not significant (P=0.9105) (Figure 19). No single serotype was associated with this increase. Overall there was no significant

overall increase or decrease in 13vPCV (p=0.9105), 15vPCV (p=0.8235) or non-vaccine serotypes (p=0.9880).

Figure 19: Invasive pneumococcal disease in Queensland children from 2007 to 2012.

Number of IPD cases is shown above each bar. The 13vPCV serotypes listed are the extra six serotypes not targeted by the 7vPCV, and the 15vPCV serotypes are the two extra serotypes not targeted by the 7vPCV and 13vPCV. Non-vaccine targeted serotypes are those not targeted by 7vPCV, 13vPCV nor 15vPCV.



There was a significantly higher number of 13vPCV serotypes isolated from younger children (0-4 years) compared to older children (p<0.0001), and significantly higher numbers of *S. pneumoniae* associated to IPD in children 0-4 years compared to older children (Figure 20 and Figure 21). Upon closer examination of these IPD in age groups, the number of 7vPCV-serotypes observed has been decreasing from 2007 through to 2011 across all age groups, however there is no observed significant difference (Figure 22). Interestingly, in the 0-4 year age group, there was a sudden decrease of 13vPCV serotypes in 2012 (n=17) compared to earlier years (average n = 30); the significance of this decrease cannot be determined at this point because we would require further time points. In contrast, the number of 13vPCV serotypes has increased in older children in 2011-2012; this may appear significant (p=0.0078) however further time points would be required for validation.

Figure 20: Children age 0-4 years have significantly higher numbers of 13vPCV serotypes associated with IPD compared to older children.

Percentage of IPD cases (%) – the proportion of 13vPCV serotypes associated with IPD in children from 2007 to 2012. One-way ANOVA was performed with cut-off p-value of <0.5 (*** p= 0.001).

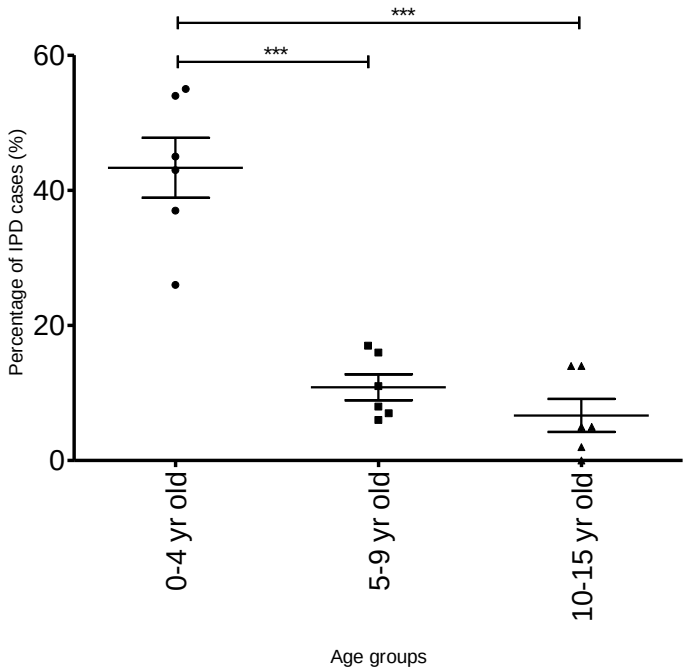
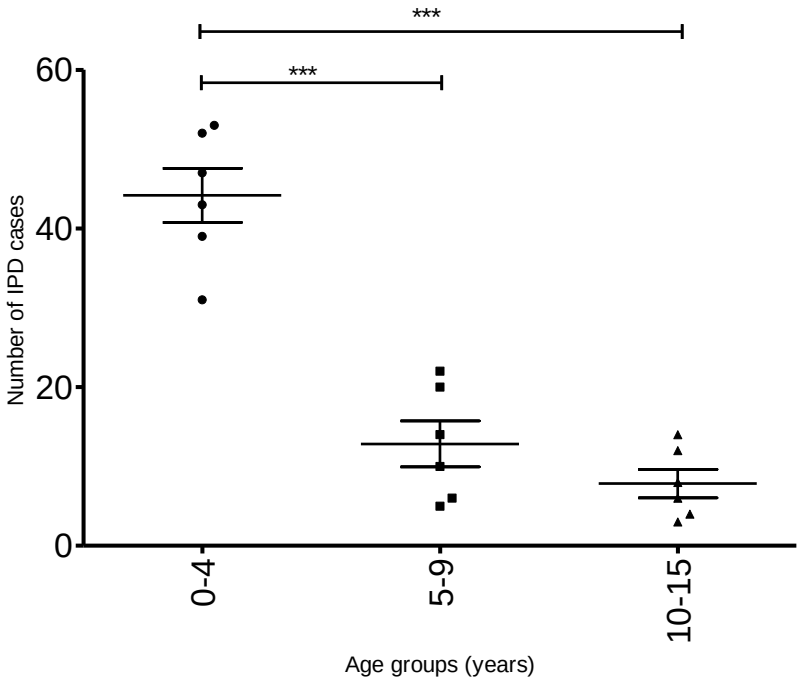


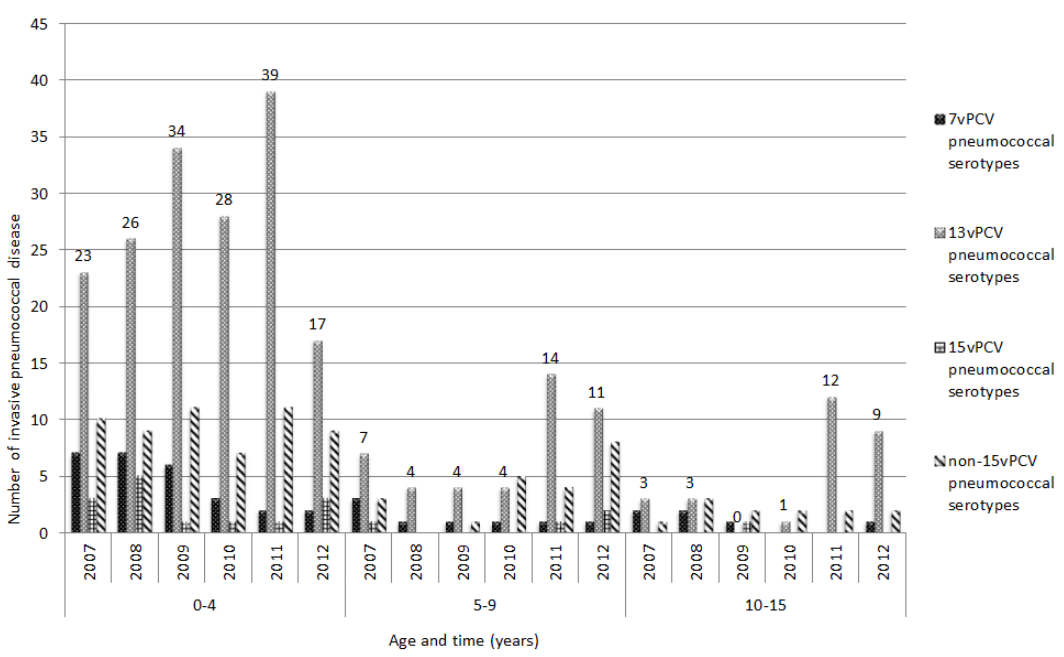
Figure 21: Children age 0-4 years have significantly higher numbers of *S. pneumoniae* associated with IPD compared to older children.



There are no significant increases in the number of 15vPCV serotypes (serotypes 22F and 33F) or non-15vPCV serotypes. However 30% of pneumococcal isolates causing IPD in 2010-2012 were caused by non-vaccine serotypes.

Figure 22: Prevalence of *S. pneumoniae* serotypes isolated from Queensland children (0-15 years) from 2007 to 2012.

The number of IPD cases is shown above bars for 13vPCV serotypes. The 7vPCV pneumococcal serotypes include serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV pneumococcal serotypes include the extra six serotypes not targeted by the 7vPCV, including serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV pneumococcal serotypes include the two extra serotypes 22F and 33F not targeted by 7vPCV or 13vPCV. Non-15vPCV pneumococcal serotypes include all those not targeted by the 7vPCV, 13vPCV or 15vPCV.

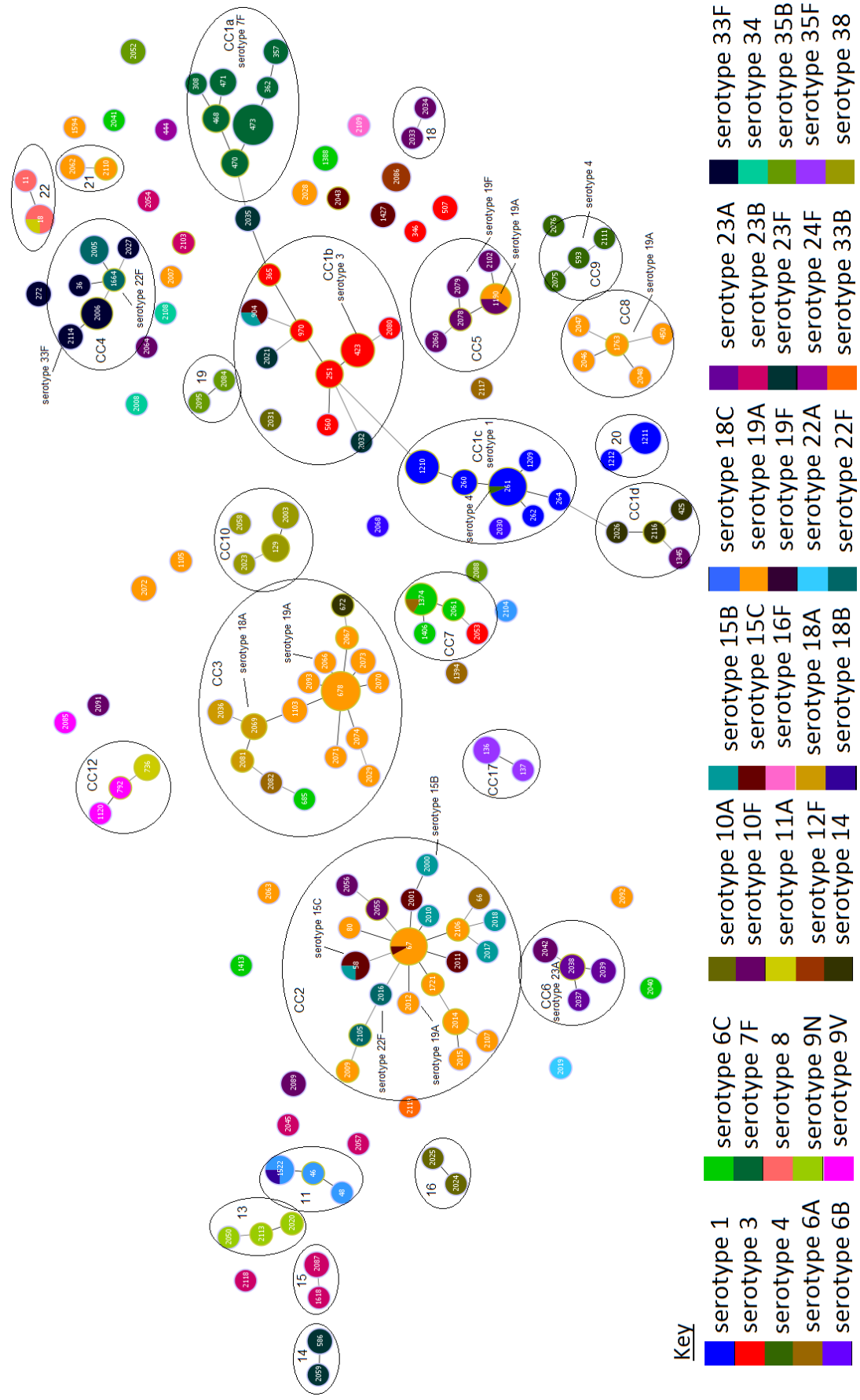


5.3.2 Pneumococcal population structure

The overall MLVA eBursts (MLVA1 and MLVA4) for 317 pneumococcal isolates demonstrates which invasive *S. pneumoniae* isolates have been circulating throughout Queensland children from 2007 to 2012 (Figure 23 and Figure 24). MLVA1 eBurst diagram indicates that there are 22 clonal complexes (CCs). There are three notable complexes - CC1 (predominantly serotypes 7F, 3 and 1), CC2 (serotypes 19A, 15B, 15C and 22F), and CC3 (serotypes 19A and 18A). CC1 appears to contain three sub-complexes of serotype 1, 3 and 7F. However, when examining the eBurst diagrams of the MLVA4 method (Figure 24), the sub-complexes disappear, and form three separate complexes; CC1 (serotype 7F), CC3 (serotype 3) and CC8 (serotype 1). In comparison to MLVA1, MLVA4 method detects a total of 32 clonal complexes, indicating a higher discrimination.

Figure 23: MLVA1 eBurst diagram with serotypes overlayed (n=317) (see opposite page).

Clonal complexes (CCs) are circled and contain single locus variants (SLV) and double locus variants (DLV). CC1 has been split into four sub-complexes (a, b, c and d). MLVA1 genotype is presented inside each circle, and size of circle represents the number of isolates detected in this study. Each colour represents a different serotype, as shown in the legend key. Circles with split colours indicate that two or more serotypes have the same MLVA1 genotype, indicating potential capsule switching.


CC1a
serotype 7F
CC1b
serotype 3
CC1c
serotype 1
CC1d
CC2
CC3
serotype 18A
serotype 19A
CC4
serotype 33F
serotype 22F
CC5
serotype 19F
serotype 19A
CC6
serotype 23A
CC7
CC8
serotype 19A
CC9
serotype 4
CC10
CC12
CC17
Key
serotype 1
serotype 3
serotype 4
serotype 6A
serotype 6B
serotype 6C
serotype 7F
serotype 8
serotype 9N
serotype 9V
serotype 10A
serotype 10F
serotype 11A
serotype 12F
serotype 14
serotype 15B
serotype 15C
serotype 16F
serotype 18A
serotype 18B
serotype 18C
serotype 19A
serotype 19F
serotype 22A
serotype 22F
serotype 23A
serotype 23B
serotype 23F
serotype 24F
serotype 33B
serotype 33F
serotype 34
serotype 35B
serotype 35F
serotype 38

MLVA4 eBurst indicates that six predominant CCs were present in Queensland during 2007-2012, and these included CC1 (serotype 7F), CC2 (serotype 19A and 19F), CC3 (serotype 3), CC5 (serotype 19A), CC6 (serotype 15B, 15C and 19A), and CC8 (serotype 1 and 4) (Figure 24). Some of these complexes are highly clonal, containing a single serotype and few MLVA4 genotypes such as CC1 (serotype 7F) and CC5 (serotype 19A), whereas the other complexes contain more than one serotype with diverse MLVA4 genotypes, such as CC6 (serotypes 19A, 15B and 15C), CC22 (serotypes 22F and 33F) and CC2 (serotypes 19A and 19F). Serotype 19F remains in circulation in Queensland despite being targeted by the 7vPCV, which was nationally funded for all infants from January 2005 to July 2011 (NCIRS, 2015). These 19F serotypes are all multi-drug resistant (unpublished data). CC6 is of interest for further examination as it also contains many strains of serotype 15C and 15B, with a potential capsule switch from a 19A to 15C.

MLST eBurst diagram indicates that at least 11 CCs are circulating throughout Queensland, however since only 202 isolates were genotyped, this is not a true representation of the population structure (Figure 25). Apart from this, all of the 13vPCV and non-13vPCV serotypes were genotyped using MLST, therefore a comparison of MLST and MLVA methods can be made. A number of MLST singletons (identified within the red box) are associated with clonal complexes when using either MLVA methods. For example, serotype 7F represents a singleton using MLST (ST191) however can be further differentiated into a clonal complex using either MLVA methods.

PLEASE SEE FOLLOWING PAGE FOR FIGURE 24.

Figure 24: MLVA4 eBurst of invasive *S. pneumoniae* isolated from 2007 to 2012 in Queensland children with serotypes overlayed (n=317) (see opposite page).

Clonal complexes (CCs) are boxed and contain single locus variants (SLV) and double locus variants (DLV). CCs linked by a thin black line indicate triple locus variants (TLV). MLVA4 genotype is presented inside each circle, and size of circle represents the number of isolates detected in this study. Each colour represents a different serotype, as shown in the legend key. Circles with split colours indicate that two or more serotypes have the same MLVA4 genotype, indicating potential capsule switching.

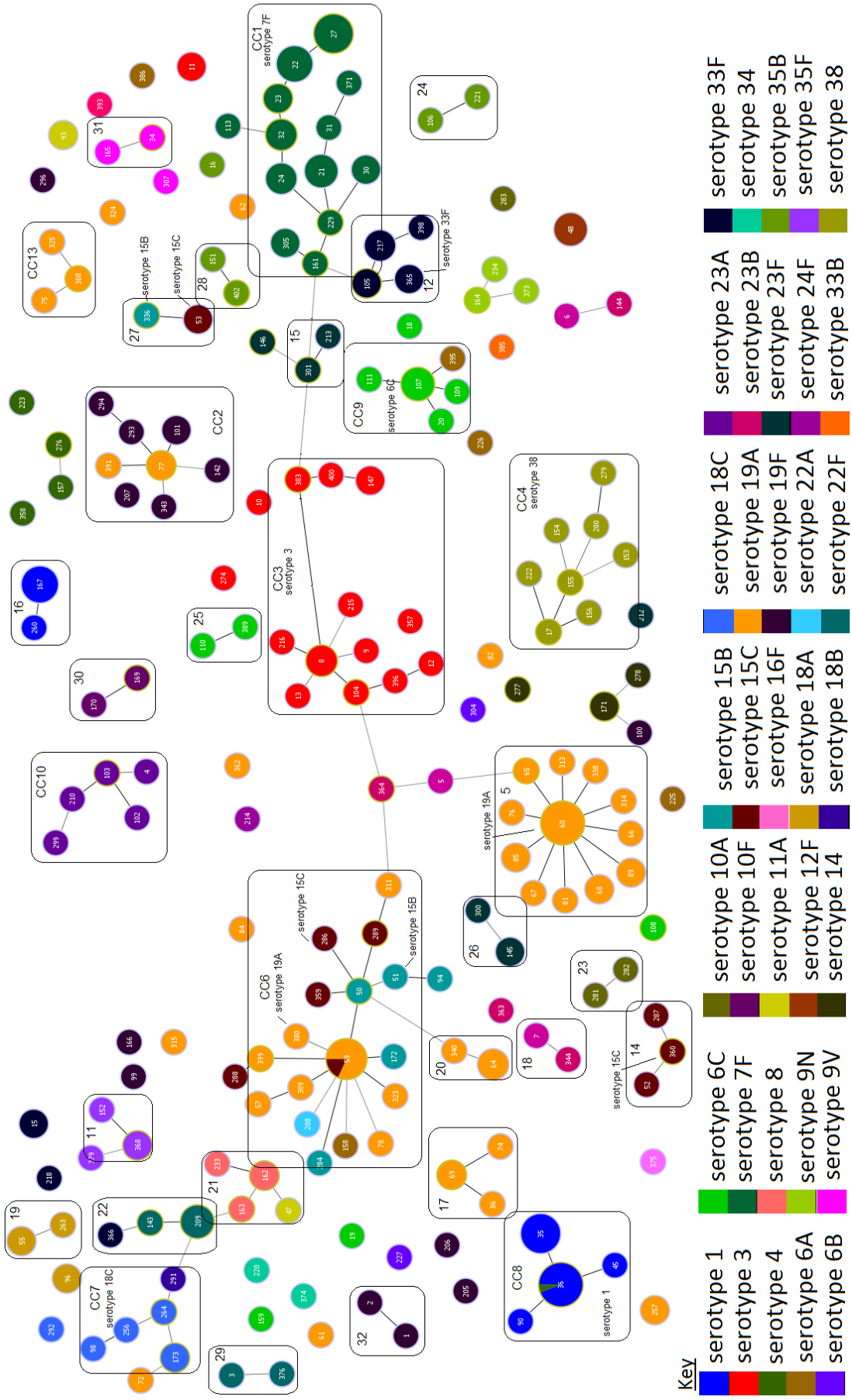
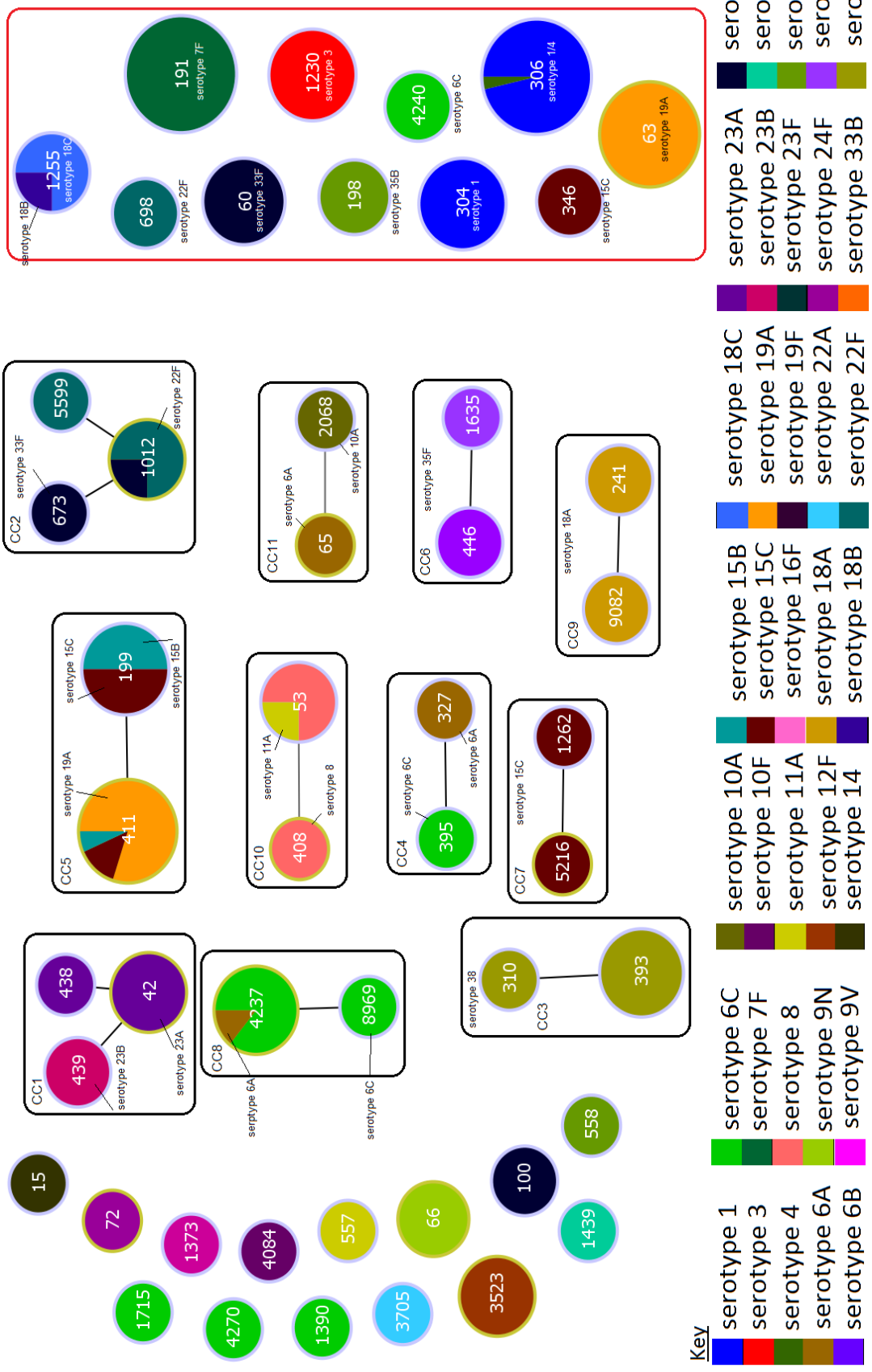

CC1
serotype 7F
CC2
CC3
serotype 3
CC4
serotype 38
CC6
serotype 19A
CC7
serotype 18C
CC8
serotype 1
CC9
serotype 6C
CC10
CC13
serotype 15B
serotype 15C
serotype 33F
Key
serotype 1
serotype 3
serotype 4
serotype 6A
serotype 6B
serotype 6C
serotype 7F
serotype 8
serotype 9N
serotype 9V
serotype 10A
serotype 10F
serotype 11A
serotype 12F
serotype 14
serotype 15B
serotype 15C
serotype 16F
serotype 18A
serotype 18B
serotype 18C
serotype 19A
serotype 19F
serotype 22A
serotype 22F
serotype 23A
serotype 23B
serotype 23F
serotype 24F
serotype 33B
serotype 33F
serotype 34
serotype 35B
serotype 35F
serotype 38

Figure 25: MLST eBurst population structure of *S. pneumoniae* in Queensland from 2007 to 2012 with serotypes overlayed (n=202) (see opposite page).

Clonal complexes (CCs) are circled and contain single locus variants (SLV) and double locus variants (DLV). MLST sequence type (ST) is presented inside each circle, and size of circle represents the number of isolates detected in this study. Each colour represents a different serotype, as shown in the legend key. Circles with split colours indicate that two or more serotypes have the same MLST genotype, indicating potential capsule switching. Circles within the red box indicate isolates that appear to be singletons (i.e. not genetically related to any other ST within this population study) when using MLST genotyping, but become CCs when using any of the MLVA methods.

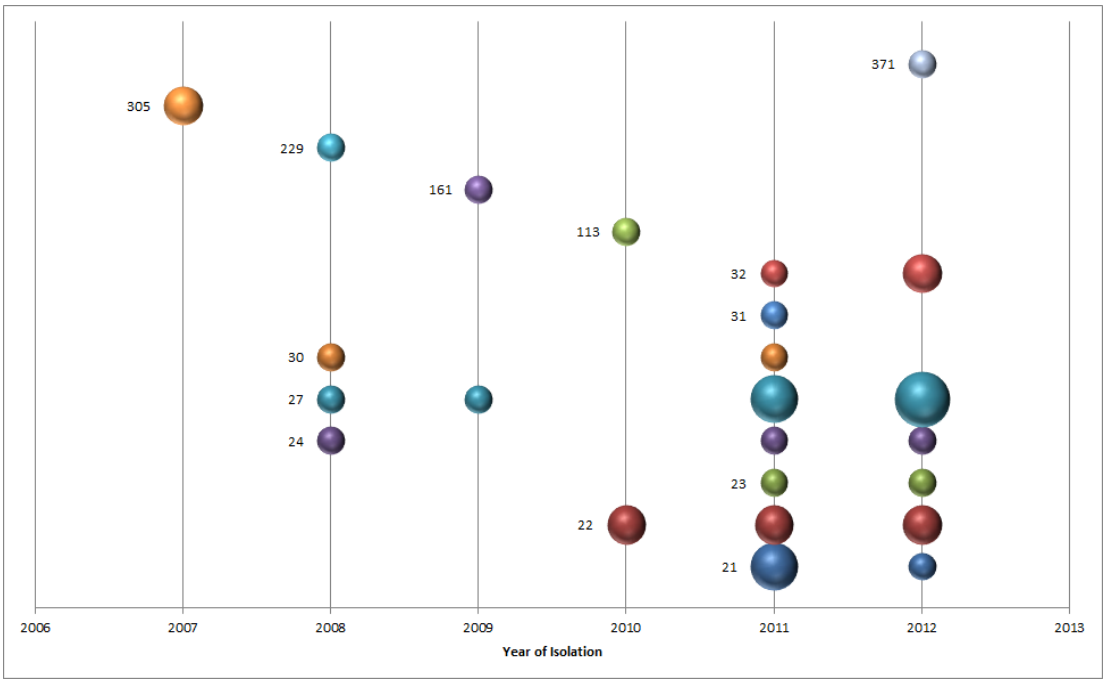

Key
serotype 1
serotype 3
serotype 4
serotype 6A
serotype 6B
serotype 6C
serotype 7F
serotype 8
serotype 9N
serotype 9V
serotype 10A
serotype 10F
serotype 11A
serotype 12F
serotype 14
serotype 15B
serotype 15C
serotype 16F
serotype 18A
serotype 18B
serotype 18C
serotype 19A
serotype 19F
serotype 22A
serotype 22F
serotype 23A
serotype 23B
serotype 23F
serotype 24F
serotype 33B
serotype 33F
serotype 34
serotype 35B
serotype 35F
serotype 38
CC1
439
438
42
serotype 23B
serotype 23A
CC2
673
5599
1012
serotype 33F
serotype 22F
CC3
310
393
serotype 38
CC4
395
327
serotype 6C
serotype 6A
CC5
411
199
serotype 19A
serotype 15C
serotype 15B
CC6
446
1635
serotype 35F
CC7
5216
1262
serotype 15C
CC8
4237
8969
serptype 6A
serotype 6C
CC9
9082
241
serotype 18A
CC10
408
53
serotype 11A
serotype 8
CC11
65
2068
serotype 6A
serotype 10A
1255
serotype 18C
serotype 18B
698
serotype 22F
191
serotype 7F
60
serotype 33F
1230
serotype 3
198
serotype 35B
304
serotype 1
4240
serotype 6C
346
serotype 15C
306
serotype 1/4
63
serotype 19A
15
72
1373
4084
1715
4270
1390
557
3705
66
3523
100
1439
558

A number of MLVA4 genotypes were more prevalent than other genotypes, including MT36 (serotype 1; n= 15), MT35 (serotype 1; n= 10), MT60 (serotype 19A; n = 10), MT59 (serotype 19A/15C; n= 8), MT27 (serotype 7F; n = 7), MT22 (serotype 7F; n= 6), MT21 (serotype7F; n= 4), and MT107 (serotype 6C; n= 4)(Figure 24).

Closer examination of some of these clonal complexes over time can help identify evolutionary trends. The highly clonal (single MLST type ST191) serotype 7F appears to be diversifying (meaning more genotypes are emerging) from 2007 to 2012 in Queensland (Figure 26), although it is unknown whether this observed change is due to diversification within this CC or whether it is because of new introductions of clones to Queensland. The variety of serotype 7F MLVA4 genotypes has increased from one (2007) to seven (2012), and the total number of isolates has been increasing from 2 isolates (2007) to 12 isolates (2012).

Figure 26: CC1 (serotype 7F) using MLVA4 genotyping from 2007 to 2012.

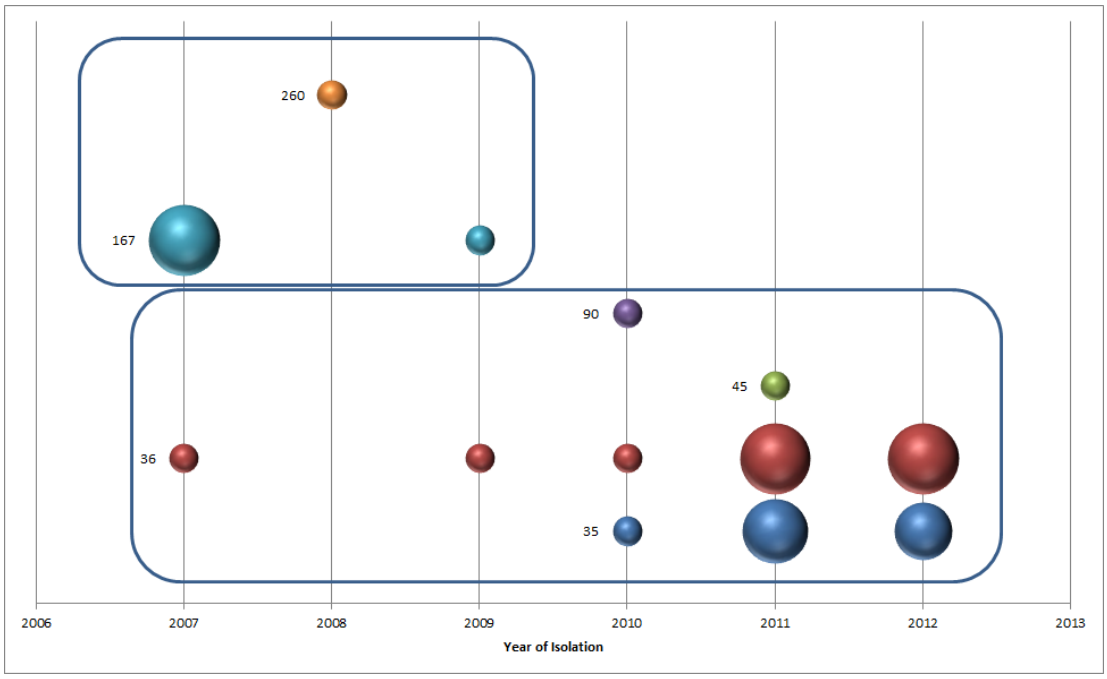
The MLVA4 genotype is shown beside the earliest detected isolate. Size of the circles indicates the number of isolates detected within each particular year. Each colour is associated to a single MLVA4 genotype.



On the other hand, serotype 1 (CC8 and CC16) trends over time seem to show that CC16 disappears in earlier years, whereas CC8 emerges and expands into several genotypes in later years (Figure 27).

Figure 27: CC8 (top box) and CC16 (bottom box) (both serotype 1) using MLVA4 genotyping from 2007 to 2012.

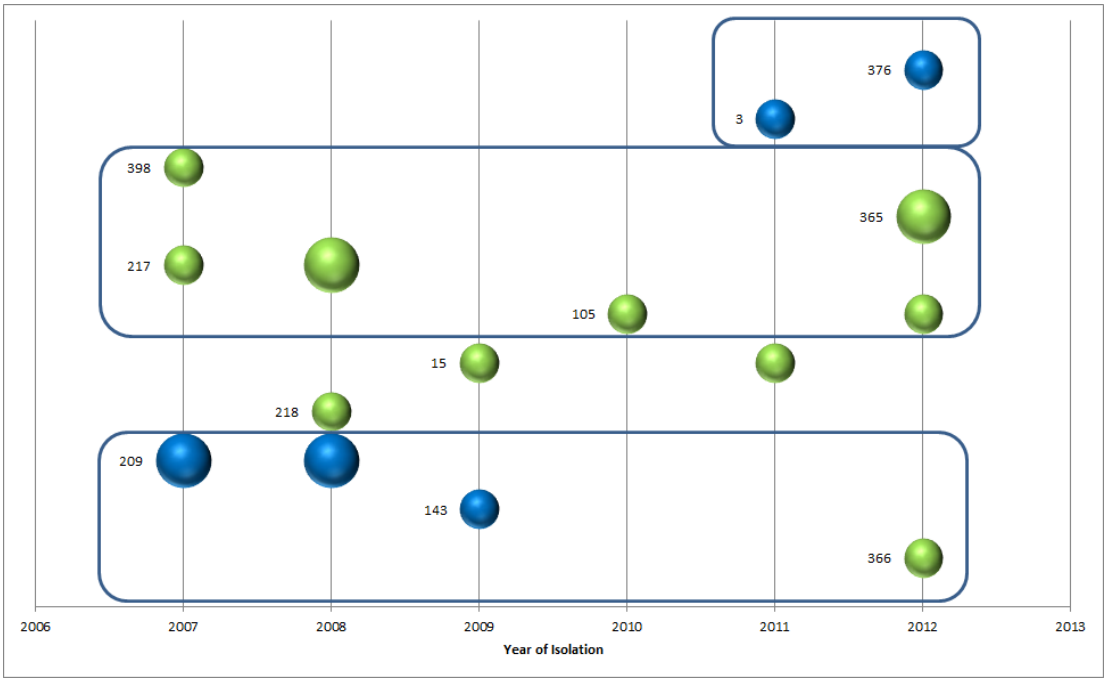
The MLVA4 genotype is shown beside the earliest detected isolate. Size of the circles indicates the number of isolates detected within each particular year. Each colour is associated to a single MLVA4 genotype.



Also of interest was whether or not there were changing trends of serotype 22F and 33F, which are the additional targets of a 15vPCV that is currently in trial to be the future pneumococcal vaccine. As shown in Figure 28, the additional 15vPCV serotypes (namely serotypes 22F and 33F) have not increased dramatically (4 isolates in 2007 compared to 5 isolates in 2012). The changing trends over time seem to corroborate with this earlier finding, as clonal complexes do not appear to be dramatically changing (Figure 28). However, we notice that a small complex of two pneumococcal MLVA4 genotypes (CC29) had emerged in recent years (2010-2012) and CC22 appears to contain both serotype 22F and 33F.

Figure 28: CC29 (top box – serotype 22F), CC12 (middle box – serotype 33F) and CC22 (bottom box - serotype 22F and 33F) using MLVA4 genotyping from 2007 to 2012.

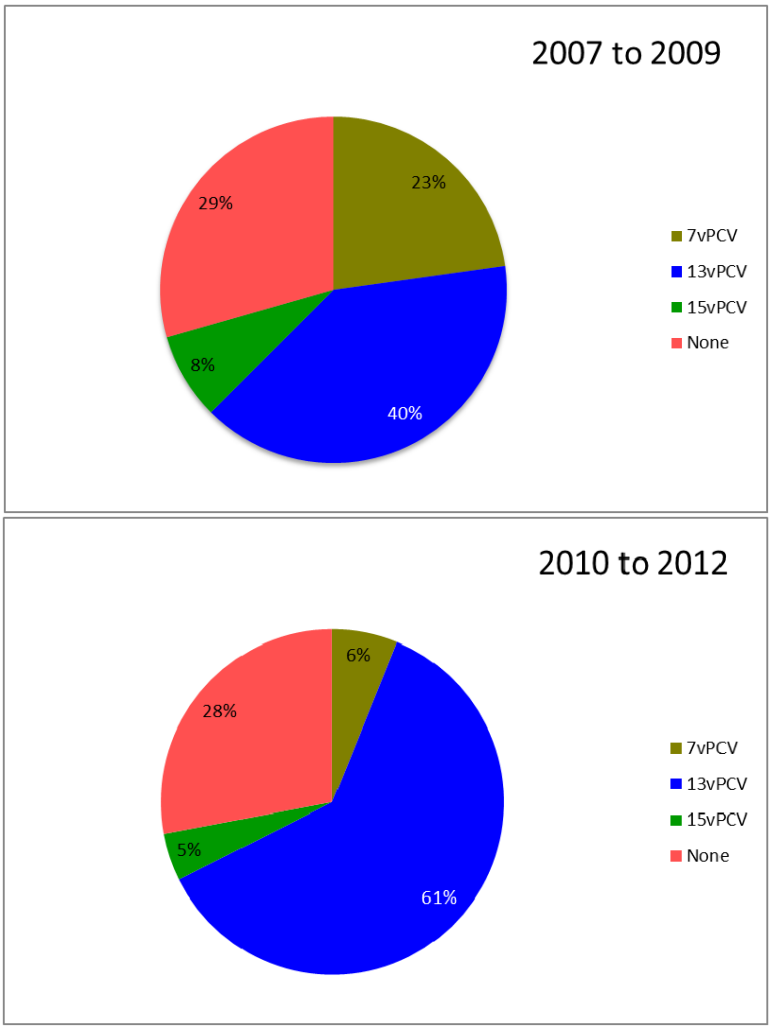
The MLVA4 genotype is shown beside the earliest detected isolate. Size of the circles indicates the number of isolates detected within each particular year. Isolates not contained by a box are not genetically related within this population study (i.e. appear as singletons). Colour represents serotype (green = serotype 33F; blue = serotype 22F).



The eBurst diagrams also allow examination of the changes in the population structure over time. Two time periods were examined; 2007 to 2009, which is post-introduction of 7vPCV in the Australian national immunisation program in 2005, and 2010-2012, which overlaps with the introduction of 13vPCV in July 2011. Comparison of these two equal time periods (3 years) enabled observations of serotype replacement between these years. The MLVA4 method shows that from 2007 to 2009, there are 17 clonal complexes identified and 40% of isolates were 13vPCV serotypes while 23% were 7vPCV serotypes (Figure 29 and Figure 30). The 7vPCV-serotypes have declined to 6% of the population however 13vPCV-serotypes have increased to 61% of the population.

Figure 29: The prevalence of Pneumococcal Conjugate Vaccine (PCV) serotypes during 2007 to 2009 compared to 2010 to 2012.

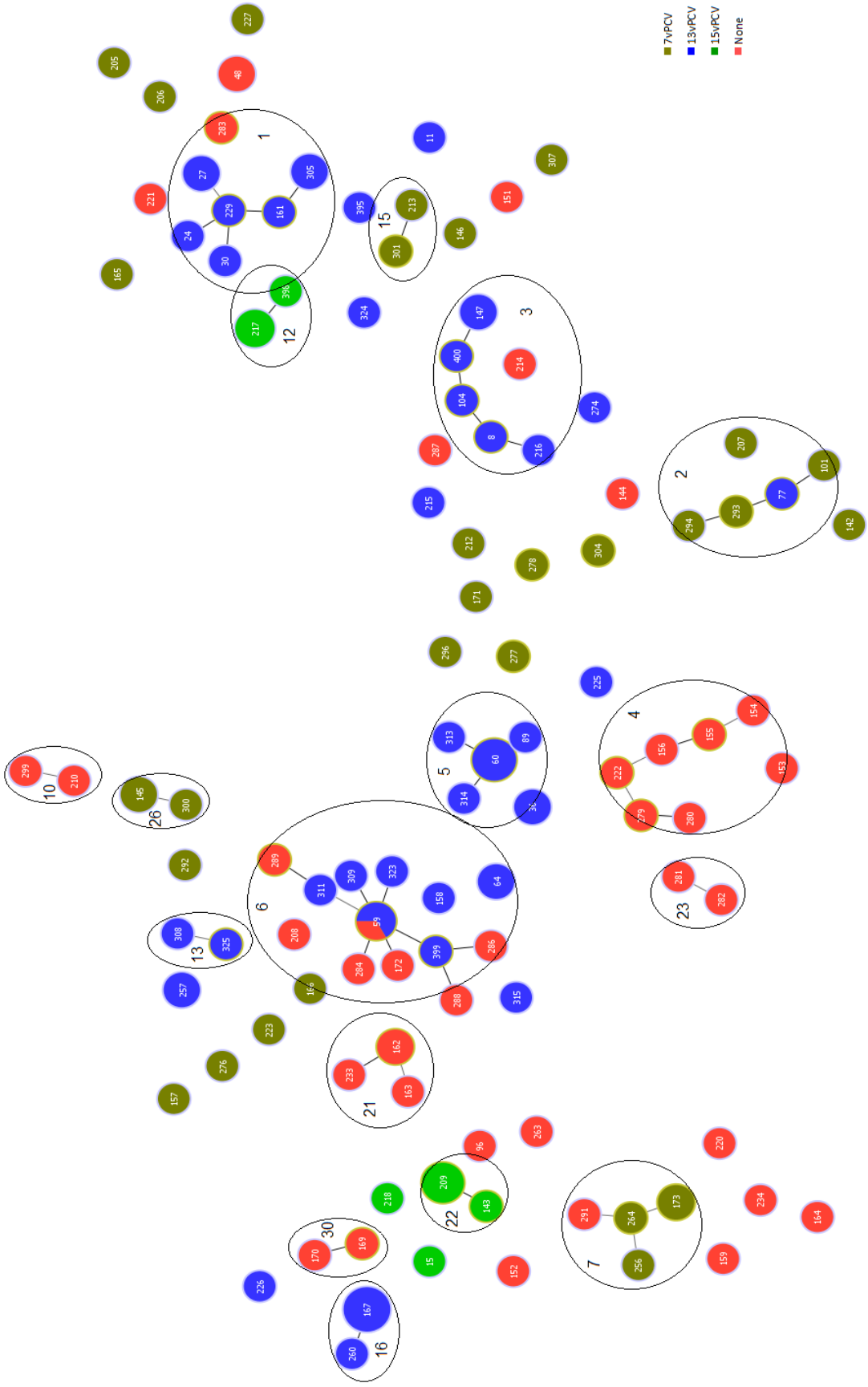
The 7vPCV serotypes include serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV serotypes include the extra serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV serotypes include serotypes 22F and 33F only. Serotypes not targeted by either of these vaccines are represented as 'none'.



There is also evidence from the current data of emerging and disappearing complexes in the Queensland pneumococcal population structure. The disappearing complexes that were present in 2007-2009 but not observed in 2010-2012 include CC4 (serotype 38), CC7 (serotypes 18B and 18C), CC13 (serotype 19A), CC15 (serotype 23F), CC16 (serotype 1), CC21 (serotypes 8), CC22 (serotype 22F), CC23 (serotype 10A), CC26 (serotype 23F) and CC30 (serotype 10F) (Figure 31). Some of these complexes contain serotypes targeted by the 7vPCV (CC7, CC15 and CC26). However, other complexes that have disappeared that were not targeted by any vaccines, including CC4, CC21 and CC30.

Figure 30: MLVA4 eBurst diagram of Queensland invasive *S. pneumoniae* from 2007 to 2009 with designated pneumococcal conjugate vaccine (PCV) serotypes (see opposite page).

Clonal complexes (CCs) are circled and contain single locus variants (SLV) and double locus variants (DLV). The 7vPCV serotypes include serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV serotypes include the extra serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV serotypes include serotypes 22F and 33F only. Serotypes not targeted by either of these vaccines are represented as 'none'.


7vPCV
13vPCV
15vPCV
None

On the other hand, CCs have emerged in 2010-2012 including CC8 (serotypes 1 and 4), CC9 (serotype 6C), CC14 (serotypes 15C), CC17 (serotype 19A), CC18 (serotype 23B), CC20 (serotype 19A), CC25 (serotype 6C), CC27 (serotype 15C), CC29 (serotype 22F), and CC32 (serotypes 19F) (Figure 31). All of these complexes (except for CC8, CC17, CC20 and CC32) contain serotypes that are not targeted by any current vaccine. Overall the 13vPCV and 15vPCV target all of the serotypes in 11 complexes (CC1, CC2, CC3, CC5, CC8, CC12, CC13, CC16, CC17, CC20, CC22, and CC29). They do not target all of the serotypes observed in 20 complex using MLVA4 (CC4, CC6, CC9, CC10, CC14, CC18, CC21, CC23, CC25, CC27, and CC30). eBurst analysis of MLVA4 data indicates that seven predominant clonal complexes were present during 2007-2012, and these include CC1 (serotype 7F), CC2 (serotype 19A and 19F), CC3 (serotype 3), CC5 (serotype 19A), CC6 (serotypes 19A, 15B and 15C), CC10 (serotype 23A) and CC12 (serotype 33F).

The comparison of the Queensland *S. pneumoniae* population in this current study against the MLST international database allows identification of genetic founders, the original genotype that is thought to have diversified (Figure 32). The comparison of the Queensland *S. pneumoniae* population in this current study against the MLVA1 online database also identifies a number of genetic founders, and although it contains fewer number of isolates, it still identifies a large clonal complex that is also observed by MLST (Figure 33). MLVA4 also identified founding genotypes in Queensland, however this is only a probable founder based on eBurst model, and the more data available the more confident about the founder assignment, therefore it is advised to use the whole database if possible. Identification of founding genotypes provides a preliminary understanding of the originating/parent genotype and hence how CC may have evolved over time.

PLEASE SEE FOLLOWING PAGE FOR FIGURE 29.

Figure 31: MLVA4 eBurst diagram of Queensland invasive S. pneumoniae from 2010 to 2012 with designated pneumococcal conjugate vaccine (PCV) serotype (see opposite page).

Clonal complexes (CCs) are circled and contain single locus variants (SLV) and double locus variants (DLV). The 7vPCV serotypes include serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV serotypes include the extra serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV serotypes include serotypes 22F and 33F only. Serotypes not targeted by either of these vaccines are represented as 'none'.

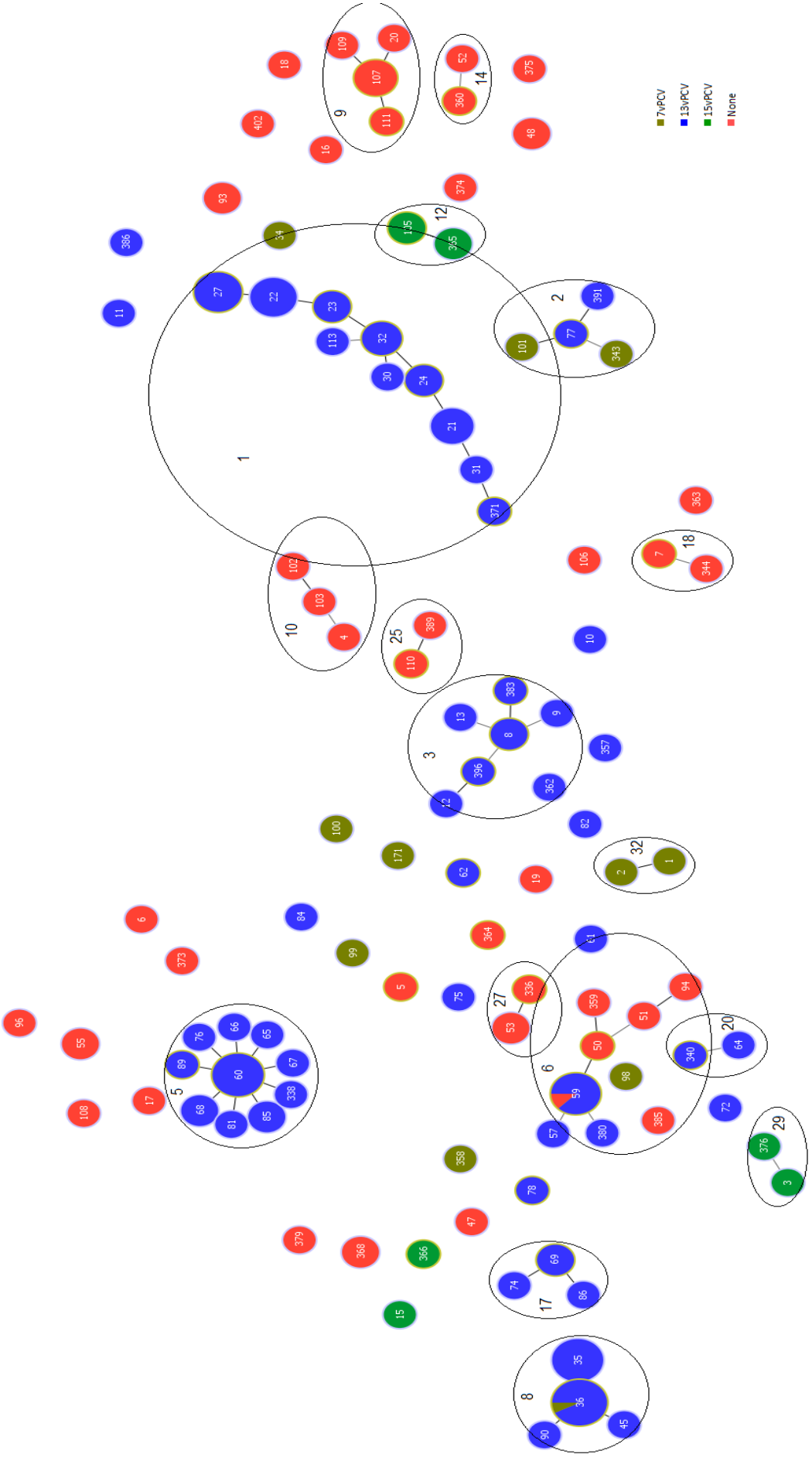

7vPCV
13vPCV
15vPCV
None

Figure 32: Queensland MLST genotypes (pink circles) compared to the *S. pneumoniae* international database.

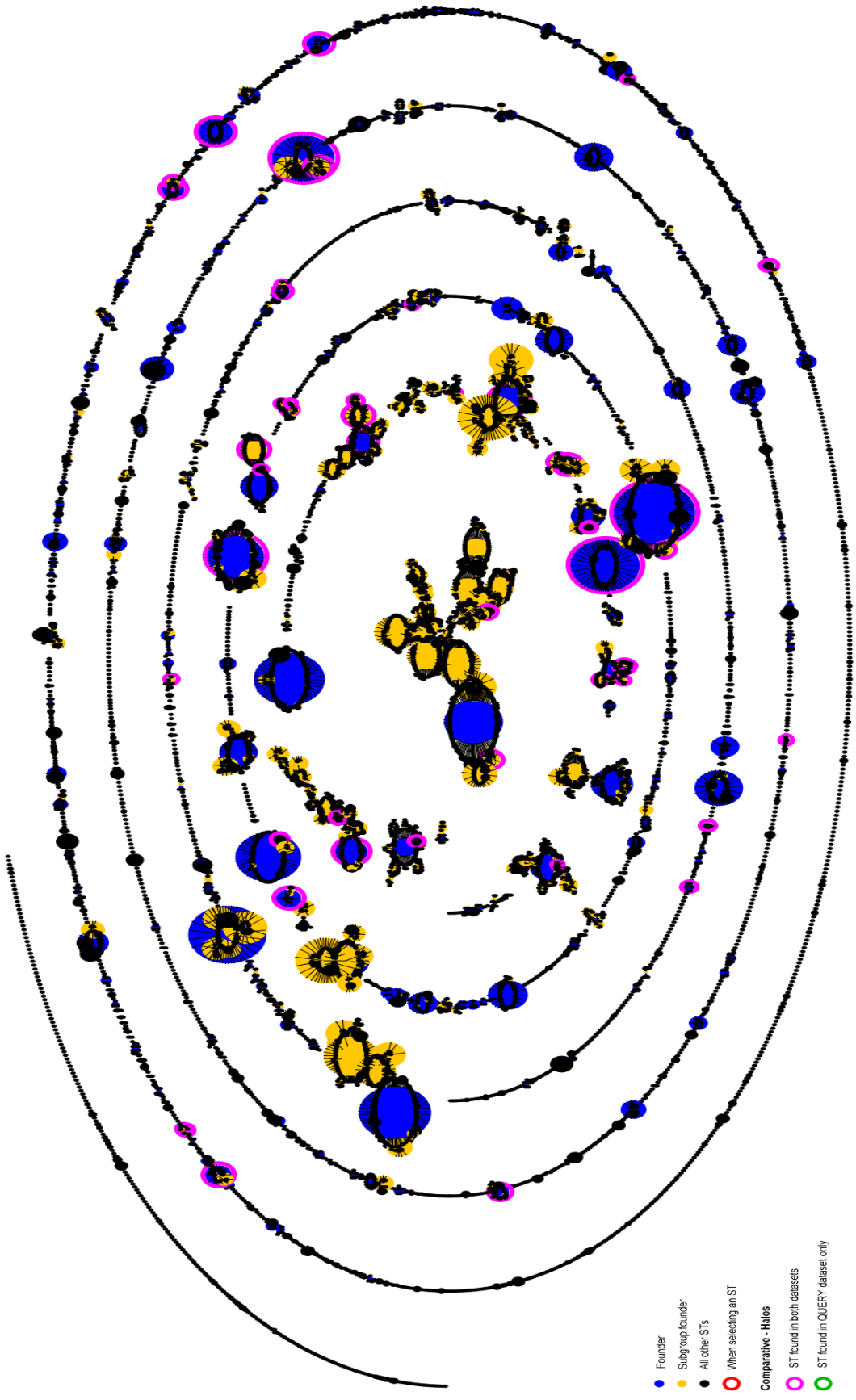
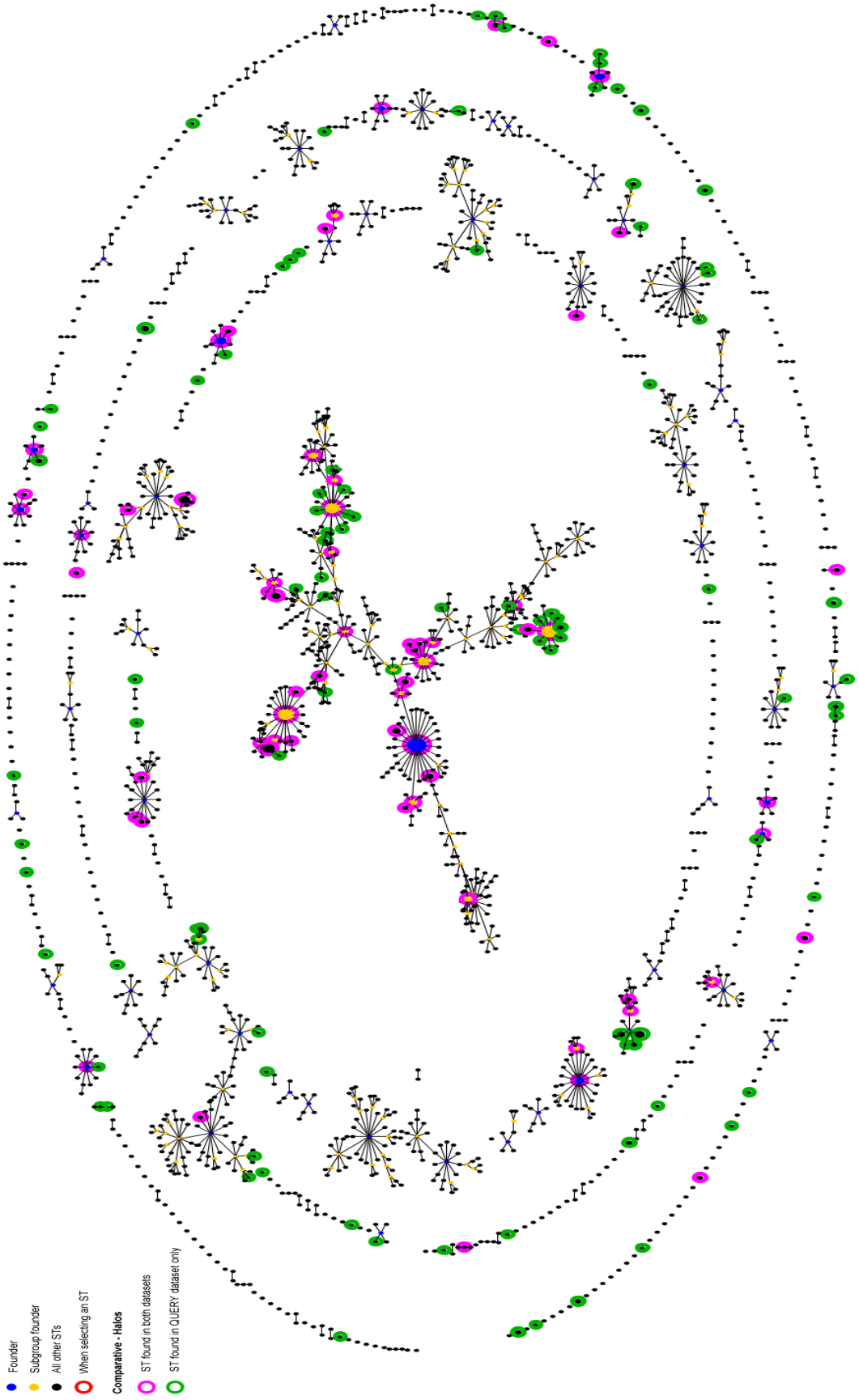


Figure 33: Queensland MLVA1 genotypes compared to the MLVA database for *S. pneumoniae*.



5.3.3 Examination of clonal complexes

Clonal complexes (CC) observed in the MLVA4 data were examined closely for increasing or decreasing clonality. A clonal complex was defined as containing MLVA4 genotypes that are single locus variants (SLV) and double locus variants (DLV). Increasing clonality is seen as a tendency towards all genotypes being single locus variants. Other observations include the number of isolates detected within each study period.

CC 1 – This clonal complex only contains serotype 7F, and with dominant MLVA4 genotypes MT21 (n=4), MT32 (n=3), MT22 (n=6) and MT27 (n=9). Serotype 7F was the second most common serotype observed, below serotype 19A. This complex has persisted from 2007 through to 2012, decreasing in clonality in 2010-2012 (i.e. becoming more diverse). The assigned founding genotype is MT32.

CC 2 – This complex contains serotype 19A and 19F, both of which are targeted by the 13vPCV. Serotype 19F is also targeted by the original 7vPCV. This complex has remained in circulation in Queensland children from 2007 through to 2012. The complex has shifted to mainly contain serotype 19A rather than 19F. This is most likely due to the serotype 19F being targeted by the 7vPCV. Despite the freely available 7vPCV implemented in 2005, serotype 19F remains in circulation. There is no founding genotype in this complex. MT101 (serotype 19F) has been observed in both 2007-2009 and the 2010-2012 periods.

CC 3 – This complex only contains serotype 3, and has persisted from 2007 through to 2012 without any major changes in the MLVA types in the clonal complex. Serotype 3 is the fourth highest IPD causing serotype in Queensland children.

CC 4 – This complex only contains serotype 38, which is not targeted by any current vaccine. This complex was dominant in earlier years (2007 – 2009).

CC 5 – This complex only contains serotype 19A, and all isolates are single locus variants of the dominant genotype MT60. This complex has also persisted from 2007-2012, and has actually decreased in clonality in 2010-2012. Of particular interest is the predicted founding genotype, MT60, which makes up most of the referred isolates in CC2 (10 pneumococcal isolates in 2010-2012). CC2 is more genetically related to CC1 (serotype 19A/15B/15C) as four MLVA alleles out of 10VNTR are the same. CC2 isolates share only two alleles with isolates in CC5 (serotypes 19A and 19F) and CC11 (serotype 19A). CC11 and CC5 are not genetically related as no alleles are common.

CC 6 - This large complex contains serotypes 15B, 15C and 19A, of which only serotype 19A is targeted by the 13vPCV. CC6 has persisted from 2007 through to 2012.

The dominant MLVA4 genotype is MT59 (n=11), with a potential capsule switch from a serotype 19A to serotype 15C. The number of MTs has declined in the 2010 to 2012 period, indicating that the complex is becoming more clonal. Serotype 19A in this complex is declining in numbers in later years, and the complex is becoming dominated by serotypes 15B and 15C. Serogroup 15 is not targeted by any childhood vaccine.

CC 8 – This clonal complex predominantly contains serotype 1 but also contains a potential capsule switch to a serotype 4 (n=1). The complex in 2010-2012 is highly clonal, and the majority of strains in this complex are MT35 and MT36. CC8 is not clonally related to CC16 (serotype 1) observed in 2007-2009. CC8 appears to have emerged in 2010-2012 either from a distantly related strain of CC16 or an independent source. Serotype 1 is the third highest serotype referred to the laboratory.

CC 9 and 25 – These complexes contain serotype 6A/6C and 6C, respectively. Both complex have emerged in 2010-2012 and are clonally unrelated to each other. Serotype 6C is not targeted by the 13vPCV. The in-trial 15vPCV does not target serotype 6C.

CC 12 – This small complex contains serotype 33F, a target of the future vaccine, 15vPCV. The MLVA4 genotypes include MT398 (n=1), MT217 (n=3), MT105 (n=2) and MT365 (n=2). CC12 has persisted from 2007 through to 2012, however it doesn't appear to be changing. Serotype 33F has been isolated more commonly from younger children (<5 years) in this study.

CC 14 and 27 – Both of these smaller complexes contain serotype 15C and 15B/15C, respectively. MLVA4 types include MT52 (n=1), MT360 (n=1), MT287 (n=1), MT336 (n=1) and MT53 (n=2). Both complexes have emerged in 2010-2012. These complexes are not clonally related to CC6 (15B/15C and 19A). No vaccine presently targets 15B or 15C.

CC 18 – This complex only contains serotype 23B which is not targeted by the 13vPCV or the in-trial 15vPCV. CC18 has emerged in 2010-2012. The total number of serotype 23B observed has increased from one (2007-2009) to six (2010-2012). There is a potential for serotype 23B to become serotype replacement, and it is not targeted by the 13vPCV and 15vPCV.

CC 22 and 29 – CC22 contains serotype 22F and 33F (serotype 33F MT366 is a single isolate in 2010-2012), whereas CC29 only contains serotype 22F. Serotype 22F (MT143, n=1) and 33F (MT366, n=1) are single locus variants, despite that serotype 22F was observed in 2007-2009 and the 33F observed in 2010-2012. CC22 is not clonally related to CC29. Both serotype 22F and 33F are targets of the in-trial 15vPCV, but are not targeted by the 13vPCV.

5.3.4 Capsule switching

We have observed capsule switches occurring in the Queensland pneumococcal population (see Chapter 4). A capsule switch is defined as two pneumococcal isolates with identical genotypes that have different capsule types/serotypes. For a more accurate depiction of whether a pneumococcus has switched its capsule, a comparison of the Queensland population against an international database is ideal.

Therefore an initial comparison of the MLST results from this study against the MLST international database indicated that several Queensland *S. pneumoniae* isolates have potentially undergone capsule switching. Similarly, comparison of our MLVA1 data with the MLVA international database revealed similar capsule switching. The MLVA4 method detected two potential capsule switches in the Queensland *S. pneumoniae* population isolates, including CC8 from a serotype 1 to serotype 4, and CC6 from a serotype 19A and 15C. Even though there is no larger database to compare the MLVA4 results with, these potential capsule switches have been observed by MLVA1 and MLST. However MLVA1 and MLST also indicate many other potential capsule switches. A study of the data generated using the MLST method indicated that there are four additional capsule switches, including serotype 22F to 33F (ST1012), serotype 18B to 18C (ST1255), serotype 15B to 15C (ST199) and serotype 8 to 11A (ST53) in the Queensland *S. pneumoniae* population. Data generated using the MLVA1 method indicated that there are five additional capsule switches, including serotype 19A to 19F (MT1190), serotype 8 to 11A (MT18), serotype 18B to 18C (MT2002), and two incidents of serotype 15B to 15C (MT58 and MT904).

A potential capsule switch from serotype 35B to serotype 7F was verified to be false in our Queensland data. It was questioned when we noticed that the ST198 was common for serotype 35B and not 7F in the MLST international database. Because serotype 7F is targeted by the 13vPCV it is hypothesised that a capsule switch would be more common to evade the vaccine (i.e. switch from serotype 7F to 35B). Natural capsule switching does occur without influence from the vaccine, however to be sure the isolate was re-serotyped, re-cultured and re-genotyped. Results indicated that the original serotype was false as the serotype 35B had been accidentally serotyped twice. A capsule switch from a serotype 35B to 7F has not been detected.

5.4 Discussion

By using MLVA4, we were able to determine the population structure of invasive *S. pneumoniae* in Queensland from 2007 to 2012. This can be used as a guide of circulating serotypes causing IPD in Queensland children five years prior to the introduction of 13vPCV in 2011. Closer examination of the Queensland pneumococcal population structure could reveal whether the pneumococcal population was changing over this time, particularly due to the vaccines 7vPCV and 13vPCV. A number of objectives were fulfilled in this study.

Firstly, we wanted to compare the use of MLST, MLVA1 and MLVA4 for determining the *S. pneumoniae* population structure in Queensland. Even though MLST was only applied to 202 isolates and MLVA applied to 317 isolates, we observed that a number of MLST singletons actually diversified and formed complexes when using either MLVA method. Higher discriminatory power can impact interpretations of population structures, especially when identifying capsule switches, clonal complexes and changes in the population. The MLVA4 method identified fewer capsule switches than the other two genotyping methods. MLVA4 genotypes may diversify within a CC, but this can make it difficult to detect the capsule switching because by definition we identify a capsule switch if two or more isolates have the same genotype but different capsules. ST types would stay consistent for longer (given the nature of stable, slow-changing house-keeping genes that are used for the ST), therefore a capsule switch would be easily identifiable because two isolates would still have the same ST even if they have different capsules. However, MLVA4 may fail to identify this because even though a capsule switch has occurred sometime during the clone's evolution, the MLVA4 type has also diversified and we detect two different MLVA4 types, which informs us that no observable capsule switch has occurred. This may be a possible limitation of MLVA4.

On the other hand, comparison of the MLVA4 population structure to patient information (not obtained in this study) may glean more accurate information or trends that MLST would fail to provide, for example since serotype 7F is known to be highly prevalent in the Australian Aboriginal communities, CC1 with a number of MLVA4 genotypes may be associated to specific communities, whereas using MLST will only determine that all serotype 7F have the same sequence type (ST191) and therefore no apparent relationships between ST191 and specific Indigenous communities.

MLVA4 also identified a number of clonal complexes, some that were only identified as singletons when using MLST. The majority of clonal complexes identified by the MLVA4 method contain a single serotype, however eight complexes contain a mix of two or more serotypes that are genetically related (Figure 24). This may imply that MLVA4 could be applied to genotype serotype by serotype. A query about whether MLVA could be used only to discriminate serotype by serotype was theorised (Van Cuyck *et al.,* 2012). It was determined that this may not be ideal in a population study because it may miss clonal relationships between different serotypes, such as those observed in this study, for example serotypes 11A and 8, serotypes 18B and 18C, serotypes 22F and 33F, and serotypes 19A, 19F, 15B and 15C. Potential capsule switching would also not be identified, for example serotype 1 and 4, and serotype 15C and 19A detected in this study. It is important to detect these relationships since serotypes 11A, 8, 18B, 22F, 33F, 15B and 15C are not targeted by the current childhood vaccines. As they are shown to be genetically similar to other serotypes, it may mean that they have the ability to switch capsules more easily or become serotype replacement strains.

The second objective was to determine the current pneumococcal population structure circulating in Queensland children. Overall IPD notification rates will not decline further with continued use of 7vPCV, therefore the introduction of the 13vPCV should provide further decrease of IPD since it targets an extra six serotypes not targeted by the 7vPCV. We have already noted an increase of 13vPCV-serotypes from 40% (2007-2009) to 61% (2010-2012) in children less than 15 years, indicating the effects of serotype replacement after the introduction of 7vPCV in 2005 (Figure 29). In South Australia, it has been reported that increases of non-7vPCV serotypes, with serotype 19A as the dominant replacement serotype, has occurred (Tran *et al.,* 2012). These authors also commented that non-7vPCV serotypes may be associated with more severe diseases, as they noticed higher rates of complications in empyema (Tran *et al.,* 2012). Oftedah *et al.* (2013) reported increases of non-7vPCV serotypes (5% to 18%) in New South Wales children before and after the 7vPCV was introduced.

An unusual spike of 13vPCV-serotypes was also detected in 2011. A serotype 1 outbreak occurred in Northern Territory, Western Australia and Queensland, which may have contributed to the unusual spike in IPD cases (Staples *et al.,* 2012; Staples *et al.,* 2015). However, even by omitting serotype 1 isolates that may be a part of an outbreak, the number of 13vPCV isolates are still unusually high. It is not known why the number of 13vPCV-serotypes has increased in 2011 (calculated to not be significant), but it appears that the introduction of the 13vPCV has come at an

appropriate time. High incidences of IPD in young children have also been reported (Roche *et al.,* 2007; Roche *et al.,* 2008; Slaon-Gardner *et al.,* 2011).

A number of CCs consisted of the same serotype as detected in other CCs, and yet these CCs were not identified to be genetically related. These included CC2 (serotype 19A/19F), CC5 (serotype 19A), CC6 (serotype 19A/15B/15C), CC13 (serotype 19A) and CC17 (serotype 19A), indicating that these five complexes may have originated from independent sources, or that the complexes have diverged from an original complex decades ago. It may be that the success of serotype 19A to cause increasing IPD cases after the introduction of the 7vPCV globally could rely on the fact that all strains are not clonally related, as demonstrated by the fact that at least five independent clonal complexes have been identified in Queensland. This could emerge as a similar issue for serotype 22F as we have identified two clonal complexes (CC22 and CC29) that contain genetically unrelated isolates of serotype 22F. Since the 13vPCV does not target 22F, it is important that these complexes are monitored. Similarly serotype 33F makes up 3% of the isolates causing IPD in children, and is not targeted by the 13vPCV. Both serotype 22F and 33F are targeted by the 15vPCV.

Serotype 7F (ST191) is highly clonal, and is the second most common serotype causing IPD in children in this study. High clonality of serotype 7F has also been observed in Scotland, Sweden and the USA (Beall *et al.,* 2006; Jefferies *et al.,* 2004; Sandgren *et al.,* 2004). It is interesting to see that serotype 7F appears to be genetically diversifying in later years, however it is unknown whether this is due to changes within the CC or whether new strains have been introduced from elsewhere into the population (Figure 26). If diversification was occurring within the CC, what pressure is there for genetic changes, and what are the potential effects on virulence with such changes? National routine typing and maintenance of a national database, which could also be compared to international data would be required to help determine the reason why certain serotypes are genetically diversifying. It is currently unknown why, but by using MLVA4 it can be shown that this 'highly clonal' serotype appears to become genetically diverse.

Interestingly, despite the introduction of 13vPCV, a number of 7vPCV-serotypes are still being observed; particularly serotype 19F which was the fifth most common serotype in this study. The 7vPCV was licensed in 2001 and made freely available for children in 2005, which means most children should have been vaccinated. There is 95% vaccine coverage in non-Indigenous citizens and 85% coverage in Indigenous citizens across Australia (Barry *et al.,* 2012). In this study we assume that all patients have been vaccinated with a pneumococcal vaccine, however due to restricted access to

patient data for this study, the vaccination history of 7vPCV is unknown. Therefore we are unable to determine whether IPD caused by serotype 19F is due to non-vaccination or vaccine failure. A 4% vaccine failure (when an appropriately vaccinated patient contracts IPD caused by a pneumococcal serotype that is targeted by the vaccine) in the Kimberley Region, Western Australia from 1995 to 2001 has been reported prior the introduction of the 23PPV (Mak *et al.,* 2004). Another possibility is that certain strains of 19F have developed the ability to evade the vaccine as it has been noted that strain MT101 (19F) has been detected in both 2007-2009 and 2010-2012. Further examination of specific strains may be required to determine whether the microorganisms have developed other mechanisms to evade the effects of the vaccine.

The third objective of this study was to determine any initial impacts of the 13vPCV on the pneumococcal population. This is one of the first published studies that reports a potential decline in 13vPCV *S. pneumoniae* serotypes associated with IPD in children 0-4 years since the introduction of the vaccine in Queensland, Australia (Figure 22), however we are unable to determine if this decline is significant as further time points are required. This initial decline is a promising result as there was a reported 168% increase in non-7vPCV serotypes in children under five in Australia in 2008 compared to 2002, with a four-fold increase due to serotype 19A (Barry *et al.,* 2012). However, it would be desirable to have at least five years surveillance data post-13vPCV introduction for a more accurate reflection on the impact of the vaccine, as recommended by Murray *et al.* (2014).

This initial decline may indicate that the 13vPCV is already starting to successfully reduce the incidence of IPD in young children, as is being observed elsewhere in the world. In the USA, since the introduction of 13vPCV in 2010, there has been a 53% decline in IPD in children under 24 months compared to the calculated average in 2007-2009. This decline was largely due to falls in 13vPCV-serotype 19A (58%), serotype 7F (54%) and serotype 3 (68%) (Kaplan *et al.,* 2013). Similarly, a decline in 13vPCV serotypes was also observed in Alaska in 2009-2011 (Singleton *et al.,* 2013). Picazo *et al.* (2013) also reported a decrease in IPD caused by 13vPCV serotypes in 2010-2011 compared to 2007-2010 in Madrid, Spain, and a study in German children has shown an overall decline in 13vPCV-serotypes causing IPD despite an increase of serotype 19A in 2010 and 2011 (Van der Linden *et al.,* 2013). A study of French children reported observing a significant lowering of carriage rates of serotype 19A, 7F and 6C with a single dose of 13vPCV, compared to children only vaccinated with 7vPCV (Cohen *et al.,* 2011), however lowering carriage rates does not necessarily reflect a decrease in invasive disease.

A steady decline in 7vPCV-serotypes causing IPD from 12 isolates (2007) to 5 isolates (2012) has also been observed, which is as expected. Serotype 1, 7F and 19A caused most cases of IPD but since the 13vPCV targets these serotypes, it is hoped further decline will occur in the future. It has also been shown that serotype 6A provides cross-protection against serotype 6C as there was a 96% OPA (opsonophagocytic assay) response (Cooper *et al.,* 2011). Since the 13vPCV only targets serotype 6A and 6B but not 6C, it is hoped that serotype 6C causing IPD will decline in the future due to cross-protection.

On the other hand this study has not observed a decline in 13vPCV serotypes in older children; in fact pneumococcal isolates increased in 2011 and 2012 (significant increase observed from 2008 to 2011; p=0.0078). This may be because children older than two years do not have access to the free Government funded vaccination schedule in Australia and are hence susceptible to the 13vPCV serotypes. However it is important to note that increased surveillance, blood culturing or even reporting changes may explain the rise in older children. It is hopeful though that as the 13vPCV is continued to be implemented for children under two, IPD cases will decline in the older age groups as vaccinated children become older. The effects of herd immunity may also decrease IPD cases in older children. Herd immunity is an important factor in reducing IPD cases in various populations (Reingold *et al.,* 2005).

The final objective of this study was to determine whether any emergence of non-13vPCV serotypes in the Queensland population could be predicted. An increase in non-13vPCV serotypes was not observed, however 33% of IPD cases in children were caused by these serotypes in 2010-2012 (Figure 29). A universal vaccine that targets all serotypes, or at least more, will most likely be required in the future.

Serotype replacement has not been observed either since the introduction of the 13vPCV in 2011 in this study, however there have been increases of serotype 6C and 23B, both not targeted by the 13vPCV. Altogether, serotype 6C, 15C and 33F (9% of isolates) could increase as they are not targeted by the 13vPCV. It is recommended that surveillance by way of MLVA continue to enable reporting of future serotype replacement. In South Australia there have also been reports of increases in non-13vPCV serotypes (Johnson *et al.,* 2012).

A number of clonal complexes have emerged and disappeared in the Queensland pneumococcal population over six years (2007 to 2012). Some clonal complexes that have disappeared include serotypes that were not targeted by any pneumococcal vaccine (CC4 –serotype 38, CC21 – serotype 8 and 11A, and CC30 – serotype 10F), indicating that the population structure can change due to natural fluctuations and are

not necessarily eliminated by vaccines. Due to some clonal complexes being highly clonal and only consisting of a single serotype, it is expected that CC1 (serotype 7F), CC3 (serotype 3) and CC8 (serotype 1) will gradually decline and disappear as they are targeted by the 13vPCV. However as serotype 1 is more commonly observed in older children, the effects of the 13vPCV may not be observed in the short term.

Conversely, a number of clonal complexes have emerged over the six years (2007 to 2012), indicating the possible beginnings of serotype replacement, especially those complexes containing serotypes not targeted by the 13vPCV. It will therefore be important to monitor the changes of CC9 (serotype 6C), CC14 (serotype 15C), CC18 (serotype 23B), CC25 (serotype 6C), CC27 (serotype 15B/15C), and CC29 (serotype 22F). Continued surveillance for changes to these clonal complexes, especially CC29, will assist in vaccine health policy decision making and the impact of introducing the 15vPCV (targets serotype 22F and 33F). Increases in serogroup 15 and serotype 22F have been reported after the introduction of the 7vPCV (Duggan, 2010; Hicks *et al.,* 2007; Jacobs *et al.,* 2008).

Furthermore, two potential capsule switches have been observed when using MLVA4. These were a switch from serotype 1 to 4, and serotype 19A to 15C. Both of these capsule switches will be verified by molecular typing method, as described in Chapter 6.

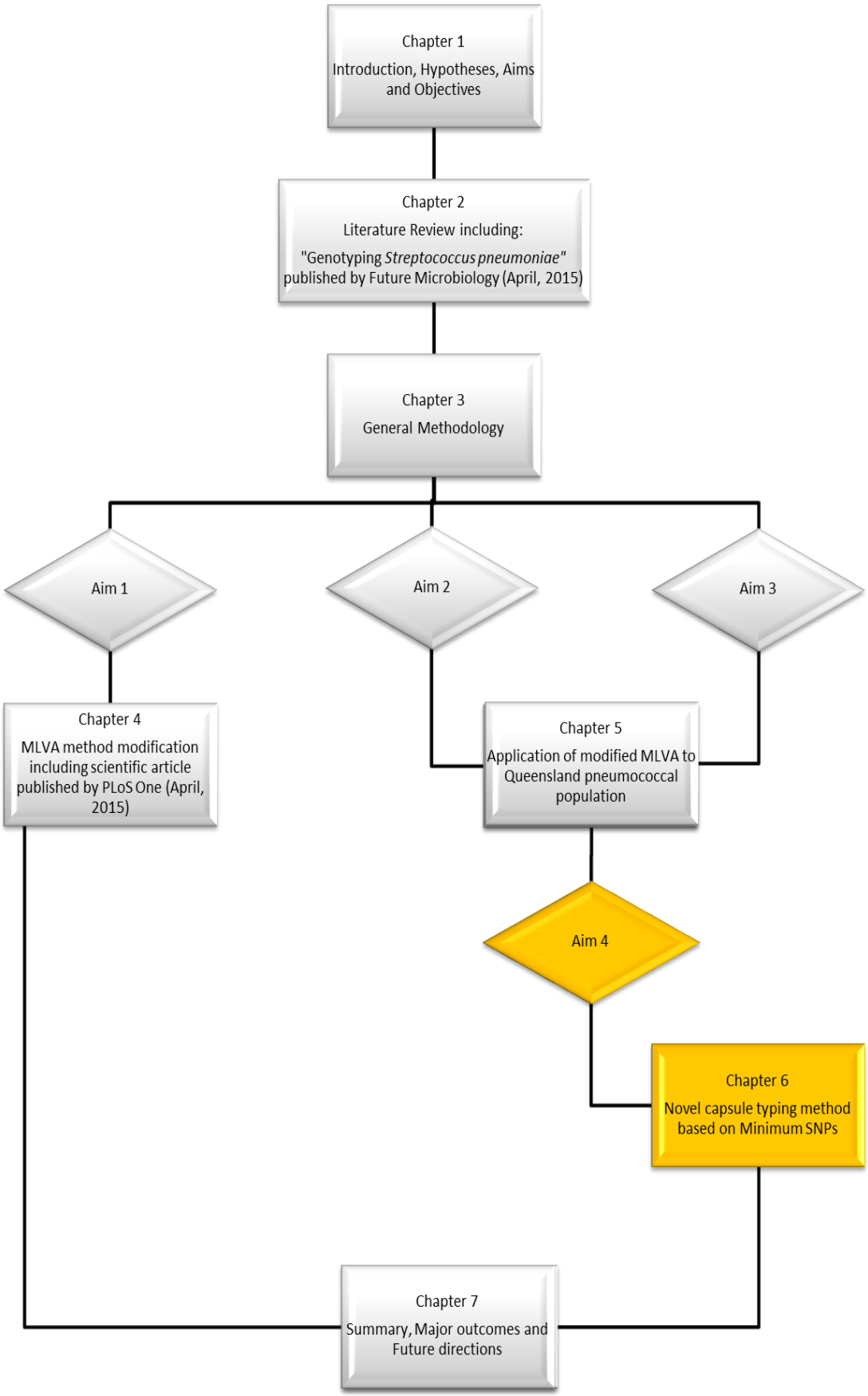
Other serotypes that deserve monitoring are serotypes 15B and 15C. Serotype 15C makes up 3% of IPD pneumococci, and could potentially become a replacement serotype. Interestingly, MLVA1 and MLST identify potential capsule switches between serotype 15B and 15C, however this can potentially be explained by the fact that serotype 15B and 15C can interconvert *in vitro* and *in vivo* at low frequencies since they only differ in the presence or absence of an acetyl group in the polysaccharide capsule (Meats *et al.,* 2003; Venkateswaran *et al.,* 1983). It would be interesting to determine whether this interconversion could aid serogroup 15 in evading any future vaccines if it did not target all serotype 15 types.

5.5 Conclusions and future perspective

Despite the introduction of 13vPCV in July 2011, preliminary results seem to indicate that this new vaccine may be already decreasing 13vPCV serotypes in young children (<5 years old) in 2012. This is a positive outcome for use of the 13vPCV. Unfortunately, an increase of 13vPCV serotypes in the older children (5-9 and 10-15 years old) was also observed since the introduction of the 13vPCV. Since these children

do not have access to the Government funded 13vPCV, it is understandable that 13vPCV-serotypes continue to cause IPD in older children. Hopefully as 13vPCV is increasingly administered to children 0-2 years, IPD cases will drop in this age group as vaccinated children become older. Also the effects of herd immunity may start to take hold, decreasing the likelihood for older siblings or children to contract IPD caused by 13vPCV-serotypes. The 2007-2012 MLVA4 genotyping data serves as a baseline for future comparisons. No changes in non-targeted serotypes were observed however they still cause about 33% of IPD in Queensland children in this study. No serotype replacement was observed, however the most dominant non-targeted serotypes that could potentially become replacements were serotypes 6B, 15C and 33F (caused 9% of IPD in this study). It will be important to continue monitoring the population structure for increases in the number of referred isolates within CC's and for expansion of clonal complexes. In this study, two potential capsule switches in *S. pneumoniae* isolates in Queensland were detected (serotype 1 to 4, and serotype 19A to 15C) using MLVA4 method but these will need to be verified by use of molecular typing methods such as sequencing since a number of other capsule switches were also detected using MLST and MLVA1.

CHAPTER 6: A NOVEL CAPSULAR TYPING METHOD FOR *STREPTOCOCCUS PNEUMONIAE* USING MINIMUM SNPS


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

6.1 Introduction

The pneumococcal capsule of the bacterium *Streptococcus pneumoniae* acts as an important colonisation and virulence factor in regards to causing human diseases. Hence this capsule has become the target of childhood vaccines 7vPCV and 13vPCV (Bratcher *et al.,* 2011; Kadioglu *et al.,* 2008). Unfortunately, the introduction of these vaccines as well as the continued use of antibiotics has resulted in a shift in the pneumococcal population structure caused by serotype replacement and capsule switching. Capsule switching was first described by Griffiths (1928). A capsule switch to a non-vaccine targeted serotype will effectively leave the pneumococci immune to the vaccine, at least until a new vaccine is developed that targets the additional serotype(s). More alarmingly, capsule switching has also led to vaccine escape strains after the pneumococcus has switched its capsule to a non-vaccine targeted capsule (previously explained in Chapter 2.10).

From the previous chapters, a number of potential capsule switches were observed in the Queensland pneumococcal population, using MLST, MLVA1 or MLVA4 combined with traditional serotyping. Using serotyping and a genotyping method (e.g. MLST or MLVA) to characterise the background genome, possible capsule switches can be detected when two pneumococci with the same genetic background have two different capsules. To verify a capsule switch, sequencing or genotyping the capsule cassette is required that ideally reveals possible points of recombination (Brueggemann *et al.,* 2007).

The capsular polysaccharide sequence (CPS) is situated between conserved regions *dexB* and *aliA* for all serotypes except for serotype 3 and 37 where the *wchE* synthase gene is defect and therefore the *tts* gene located elsewhere in the genome is active (Bentley *et al.,* 2006; Jiang *et al.,* 2001; Wyres *et al.,* 2012). *DexB* and *aliA* do not participate in capsule formation (Jiang *et al.,* 2001). Sequence analysis of 88 CPS loci reveals variable lengths (10 – 30kb) and a range of genes (Wyres *et al.,* 2012). The CPS consists of homologous regions at 5' end (*wzg, wzh, wzd* and *wze*) followed by *wchA* that encodes for the initial glucose phosphate transferase that most likely aids recombination (Bentley *et al.,* 2006; Jiang *et al.,* 2001). The *S. pneumoniae* serogroups are broken into respective serotypes (e.g. serogroup 6 has eight serotypes 6A, 6B, 6C, 6D, 6E, 6F, 6G and 6H) (Table 19). Some serogroups do not have serotypes, such as serogroup 3, and therefore is designated a serotype as itself (e.g. serotype 3).

Table 19: The genetic differences between *S. pneumoniae* serogroups/serotypes based on the capsular locus.

Some serogroups do not contain distinct serotypes, and the "Serotypes" column is therefore left blank. Capsular differentiation has not been provided for all serogroups because there is lacking information in the literature.

Serogroup	Serotypes	Capsule differentiation	Reference
1			Bentley *et al.,* 2006
2			Bentley *et al.,* 2006
3		Serotype similar to serogroup 25 and serotype 38. *wchE* synthase gene defect.	Bentley *et al.* 2006; Mavroidi *et al.,* 2007
4			Bentley *et al.,* 2006
5			Bentley *et al.,* 2006
6	A, B, C, D, E, F, G, H	6A and 6B differ at *wciP* by single base pair at positions 192, 195 or 254. 6A, 6C and 6D differ at *wciN* locus, which is also longer in 6A than 6C. 6C has unique *wciN*. 6D has added $wciP_6$. 6H is a hybrid of 6A and 6B.	Mavroidi *et al.,* 2004; Bentley *et al.,* 2006; Bratcher *et al.,* 2011; Calix *et al.,* 2010; Elberse *et al.,* 2011c; Park *et al.,* 2007; Oliver *et al.,* 2013; Park *et al.,* 2015
7	A, B, C, F	Serogroup splits into two syntenic groups – 7A and 7F have *wcwC* whereas 7B and 7C don't. A frame shift mutation in *wcwD* in 7A distinguishes it from 7F. 7B and 7C are very similar to serotype 40.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
8			Bentley *et al.,* 2006
9	A, L, N, V	9A and 9V differ from 9L and 9N by insertion of *wcjD* in former pair. Another gene *wcjE* is functional in 9V but mutated in 9A, 9L and 9N.	Bentley *et al.,* 2006
10	A, B, C, F	Serogroup splits into two syntenic groups – a present *wciG* in 10C and 10F differentiate them from 10A and 10B	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
11	A, B, C, D E, F	Serogroup splits into two syntenic groups – 11A/11D/11F and 11B/11C. Disruption of *wcjE* in 11E distinguishes it from 11A. A frame shift in *gct* in 11F and 11B differ them from 11A and 11C. 11E has disrupted *wcjE*.	Bentley *et al.,* 2006; Calix & Nahm, 2010; Mavroidi *et al.,* 2007
12	A, B, F	Nucleotide difference in *wciI* and *wcxB* separates 12A from 12F. The capsule cassette is very similar to serotype 44 and 46	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
13			Bentley *et al.,* 2006
14		Similar capsule cassette to serogroup 15, except 3' end differs with replacement of *lrp* in serotype 14 instead of glycerol-2-phosphate related genes.	Bentley *et al.,* 2006
15	A, B, C, F	Linear polymer in 15A and 15F but a branched polymer in 15B and 15C. Variable length TA repeats at 5' end of *wciZ* differentiates 15B from 15C. Similar capsule cassette to serotype 14.	Bentley *et al.,* 2006; Jones & Lemercinier, 2005
16	A, F		Bentley *et al.,* 2006
17	A, F		Bentley *et al.,*2006

Serogroup	Serotypes	Capsule differentiation	Reference
18	A, B, C, F	Two base pair difference between 18B and 18C. Frame shift mutation of *wciX* in 18B but not 18C.	Bentley *et al.,* 2006Mavroidi *et al.,* 2007
19	A, B, C, F	Serotype 19A and 19F differ from 19B and 19C by an extra sugar molecule and disaccharide side chain. A difference in polymerisation linkage and different *wzy* differ 19A from 19F.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007; Morona *et al.,* 1999a
20	A, B	There is 99% similarity between 20A and 20B. A mutation in *whaF* in 20A differs from 20B.	Calix *et al.,* 2012
21			Bentley *et al.,* 2006
22	A, F	Serotype 22A and 22F are syntenic	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
23	A, B, F	All three serotypes are similar. A *wzy*-19 is present in 23B and 23F, whereas *wzy*-30 is present in 23A.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007; Morona *et al.,* 1999b
24	A, B, F	An extra *rbsF* is present in 24B and 24F but not 24A. Genes *abp1* and *rmlC* are frame shifted in 24B.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
25	A, F	Both serotypes similar to serotype 3 and 38.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
27			Bentley *et al.,* 2006
28	A, F	Both serotypes form a sub-complex.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
29			Bentley *et al.,* 2006
31			Bentley *et al.,* 2006
32	A, F	Both serotypes form a sub-complex. 32A has a sequence difference in *wcyH* leading to an extra acetyl group at ∝-D-Glcp.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
33	A, B, C, D, F	Serogroup splits into two syntenic pairs – 33A/33F and 33B/33D. 33A contains intact *wcjE* whereas 33F is disrupted	Bentley *et al.,*2006; Calix & Nahm, 2010; Calix *et al.,* 2012; Mavroidi *et al.,* 2007
34			Bentley *et al.,* 2006
35	A, B, C, F	Serogroup 35 similar to serotype 42. A defective transposase and frame shift mutation in *wcrK* differentiates 35A fr	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
36			Bentley *et al.,* 2006
37		*wchE* synthase gene defect.	Bentley *et al.,* 2006
38		Serotype similar to serogroup 25 and serotype 3. Serotype 38 differs by presence of *wcyV* instead of *wcyE*.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
39		Contains a *wcrC*-related gene that closely resembles *wefM* of *Streptococcus oralis*.	Bentley *et al.,* 2006; Bush *et al.,* 2014
40			Bentley *et al.,* 2006
41	A, F	Serotype 41A has 6 carbohydrate units in a doubly-branched repeating unit. Acetyltransferase acts on the rhamnopyranosyl sidechain E in serotype 41F but not 41A.	Bentley *et al.,* 2006; Peterson *et al.,* 2014
42		Serotype 42 is similar to serogroup 35.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
43			Bentley *et al.,* 2006
44		A frame shift of *fnlC* is present. Very similar capsule cassette to serogroup 12 and serotype 46.	Bentley *et al.,* 2006; Mavroidi *et al.,* 2007
45			Bentley *et al.,* 2006

Serogroup	Serotypes	Capsule differentiation	Reference
46		Very similar capsule cassette to serogroup 12 and serotype 44.	Bentley *et al.,* 2006
47	A, F		Bentley *et al.,* 2006
48			Bentley *et al.,* 2006

Besides from vaccine immunity and vaccine escape, capsule switching events can also have alarming impacts on pneumococcal virulence. Just to reiterate from Chapter 2.10, a study demonstrated that capsule switching can enhance *S. pneumoniae* serotype 3 virulence in respiratory diseases (Sabharwal *et al.,* 2014). Since serotype 3 is common in Australian populations, it would be interesting to discover whether other serotype switches have an effect on virulence. The fact that the background genome may impact the virulence properties of *S. pneumoniae* in combination with the capsule cassette is interesting. It has been hypothesised that BOX elements used in several MLVA methods may have regulatory functions associated with virulence as they are located near such genes (Kolkman *et al.,* 1997; Martin *et al.,* 1992).

Since it is important to accurately detect pneumococcal capsules, several capsule typing methods have been developed. The traditional serotyping method (Quellung reaction) is universally used and considered the 'gold standard'; however it is highly expensive, limiting its use to large research laboratories, requires pure culture capsulated pneumococci and requires multiple Factor serum to distinguish between some serotypes.

On the other hand, complete analysis of the capsule cassette would undoubtedly rely on whole sequencing. Sequencers such as Illumina Genome Analyzer or Illumina HiSeq have been used (Croucher *et al.,* 2012). Most of the capsule cassettes were determined by using Expand Long Template PCR System from Roche which is able to sequence up to 20kB (Ansaldi *et al.,* 2011; Bentley *et al.,* 2006; Brueggemann *et al.,* 2007; Wyres *et al.,* 2013). Expand Long Template PCR system with DNAse I digestion and sequencing on AB377A has also been used (Jiang *et al.,* 2001). While whole CPS sequencing would produce complete results, the cost of applying this for routine diagnosis would be currently too expensive; therefore cheaper and faster PCR techniques have been developed to determine capsule types.

Multiplex PCR has also been adopted for sequencing of short regions within the capsule cassette. Pai *et al.* (2006) developed a seven multiplex PCR with 29 primer pairs which could identify 17 serotypes. This method has been widely applied and optimised to regional pneumococcal populations around the world. Ahn *et al.* (2012) modified this multiplex PCR to a 35 primer pair; however both methods still utilise gel electrophoresis for analysis.

Bratcher *et al.* (2011) sequenced the *wciP, wzy* and *wzx* genes for a capsule profiling technique for serogroup 6. The *wciN* gene and flanking regions were sequenced to analyse evidence of recombination. However, as *wzy* and *wzx* are highly conserved across all serotypes, this profiling method would not be suitable as a capsule typing method that distinguishes all serotypes.

Similarly, Elberse *et al.* (2011c) used a long-PCR method and allele-specific PCR using *wzy* and *rmlC* to characterise serotypes 19A and 19F, and serogroup 6. This method was specific enough to determine capsular subtypes within each serotype that could allow further information about microevolution. The significance of determining capsule sub-types may reveal a genetic relationship between the emergence or success of certain *S. pneumoniae* strains and the types of vaccines used. Elberse *et al.* (2011c) identifies that serotype 19A subtype II (19A-II) may have enhanced polysaccharide production due to a difference in the *wzg* locus compared to serotype 19A-I, particularly since subtype 19A-II has been observed more frequently in vaccinated children. Similarly, Mavroidi *et al.* (2004) devised a CPS profiling method for characterising serogroup 6, produced from the presence or absence of 254 homology groups within the genome. In this study homology groups were defined as cps genes that shared significant sequence homology using the program TribeMCL (Mavroidi *et al.,* 2004). Other PCR based serotyping, as previously discussed in Chapter 2 Literature Review are presented in *Appendix A4: Comparison of various pneumococcal PCR-based serotyping methods.* While most of these capsule typing methods have enhanced knowledge of specific serotypes, there is no current PCR based method that characterises all 98 known serotypes.

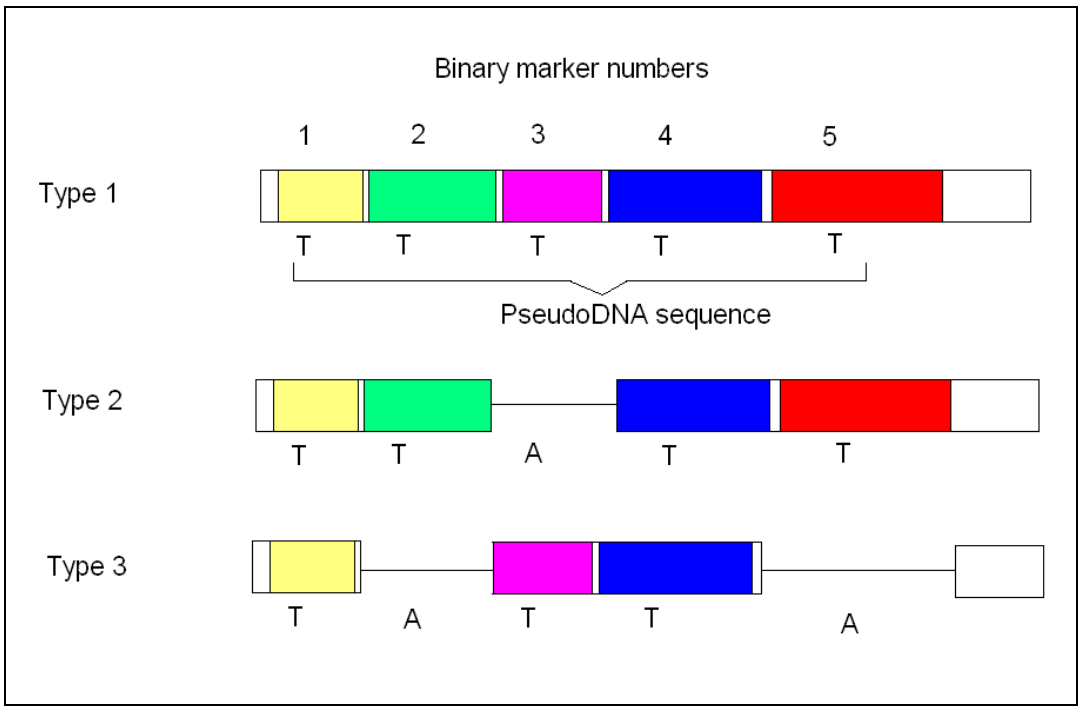
This study examines a novel systematic approach for the identification of suitable genetic targets to efficiently genotype the capsule locus. A bioinformatics program called Minimum SNPs has been used for allele-specific characterisation of certain bacteria by selecting highly discriminatory targets or single nucleotide polymorphisms (SNPs) within a genome with the maximum Simpson's Index of Diversity (D) (Robertson *et al.,* 2004; Stephens *et al.,* 2007). Minimum SNPs has been applied to genotype or characterise *Neisseria meningitidis, Enterococcus faecalis, Enterococcus faecium, Escherichia coli, Campylobacter jejuni* and *Staphylococcus aureus* as an alternative method to MLST (Price *et al.,* 2006; Rathnayake *et al.* 2011; Sheludchenko *et al.,* 2010; Stephens *et al.,* 2007). Stephens *et al.* (2007) applied this Minimum SNPs to characterise *mec* cassette in *S. aureus*, by identifying 22 out of 34 target genes to genotype 46 Staphylococcal cassette chromosomal *mec* (SCCmec). The algorithm makes

no assumptions concerning the population structure or diversity (Robertson *et al.,* 2004).

This similar idea can be applied to the *S. pneumoniae* cassette-like capsule locus to characterise the serotypes. This assay can be used to identify a highly discriminatory minimum set of genes (instead of the traditional SNPs) within the CPS cassette that characterises all or most of the serotypes (Robertson *et al.,* 2004). The bioinformatics computer program Minimum SNPs has been developed to "identify highly informative sets of SNPs" by using a binary system of T = present and A = absent to identify each binary marker (Figure 34) (Robertson *et al.,* 2004). A pseudoDNA sequence is produced for each capsule type and Minimum SNPs analyses the combinations of binary markers to maximise the Simpson's index diversity (D). A minimum set of 'targets' (capsule genes) are identified with the highest D-value and are used to characterise clinical strains of bacteria. In the example shown, binary markers 2 and 3 provide a D=1 i.e. they completely differentiate between each serotype (Figure 34). As a result, only binary markers 2 and 3 need to be amplified using PCR, and this is called the pneumococcal 'barcode'.

Figure 34: An illustration of the Minimum SNPs strategy used to identify the minimum 'targets' to characterise each capsule type (3 types listed).

To differentiate between 3 strain types (example only), the Minimum SNPs program selects binary marker 2 and 3 which provides the highest Simpson's Index of Diversity (D=1). Binary marker 5 could also be selected to replace binary marker 2. Therefore, amplification of binary marker 2 (e.g. using PCR) should be present in strain type 1 and 2, but absent in strain type 3. Similarly, binary marker 3 should be present in strain type 1 and 3, but absent in strain type 2. T= present genes; A= absent gene; pseudoDNA sequence is the combination of T's and A's.



Primers will be designed for the selected genes (e.g. binary marker 2 and 3 from Figure 34), and conventional or real-time PCR will be used to detect the presence or absence of genes that determine pneumococcal capsular types. A present gene is identified by an amplified band observed using gel electrophoresis (conventional PCR) or a melt curve using real-time PCR; on the other hand, an absent gene is identified by observing no visible band on gel electrophoresis or no visible melt curve. The barcode for each pneumococcal strain is then formed (will vary in length depending on the number of target genes required). Allele-specific PCR (e.g. using SNPs within the target genes) will be used to distinguish subtypes, if applicable – this would be used if two pneumococcal strains appear to have the same barcode associated to a known serogroup and we know that the serogroup can be split further into serotypes. Furthermore, capsule switching can be verified by combining the Minimum SNPs assay with MLVA genotyping.

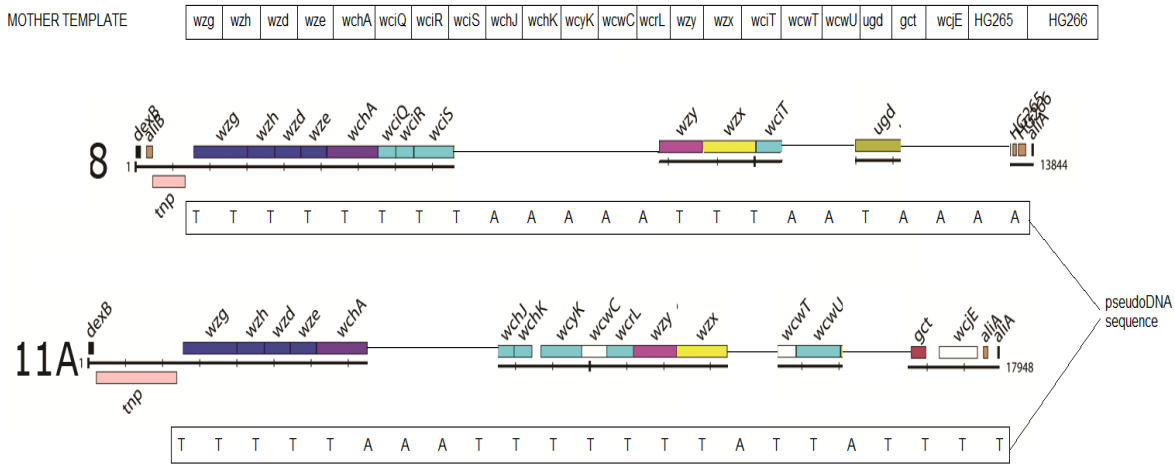
Therefore to summarise, this study aims to develop a novel capsule typing method based on the bioinformatics bi-numeral Minimum SNPs assay to characterise all 98 known pneumococci capsule genes. The objective is to improve on current molecular capsule typing methods. By using gene-specific real-time PCR and allele-specific real-time PCR, Queensland isolates of invasive *S. pneumoniae* will be tested. This capsule typing method would also be used to determine capsule switching in invasive *S. pneumoniae* in conjunction with our previously developed MLVA4 genotyping method. It is hypothesised that the novel capsule typing can detect capsule types and sub-types of *S. pneumoniae* and that capsule switching events can be verified by combination of the novel capsule typing with MLVA4. This will be a novel application of the Minimum SNPs to *S. pneumoniae* to create real-time PCR and allele-specific PCR for characterising serotypes, capsule sub-types and capsule switches in conjunction with a genotyping method.

6.2 Method

6.2.1 In silico *analysis and MinimumSNPs*

A search of the NCBI database and literature for all known sequenced *S. pneumoniae* capsule cassettes was performed (Bentley *et al.,* 2006; Bratcher *et al.,* 2011; Calix *et al.,* 2012; Park *et al.,* 2007;). A "mother template" consisting of approximately 182 possible capsular genes that have been found in the available pneumococcal capsule sequences was manually created (*Appendix A5:* In silico *analysis of 182 targets in the pneumococcal capsule cassette)*. For each of the known serotypes, an artificial binary genetic code called a "pseudoDNA" sequence was created based on T = present, and A = absent (example shown in Figure 35). An allele code for each serotype was designated to each pseudoDNA sequence (e.g. serotype 1 = cps_1, etc.).

Figure 35: An example of a 'pseudoDNA' sequence for *S. pneumoniae* serotype 8 and 11A based on the binary system of T = present and A = absent for each capsule gene that makes up the 'mother template'.



All *S. pneumoniae* pseudoDNA sequences were analysed by Minimum SNPs program following the software protocol (Stephens *et al.,* 2007). The program determined the minimum number of targets (CPS genes) that would provide the maximum Simpson's Index of Diversity (D=1.0) for all known *S. pneumoniae* serotypes. Using the chosen set of target genes, primers were designed using PrimerExpress v2.0 software (Applied Biosystems, Australia). A Clustal Omega Alignment was performed to ensure that the primers would bind across all expected serotypes. If SNP differences were found in the primer binding sites, degenerative primers were designed (e.g. 50% have alanine (A) and 50% have a cytosine (C) at the same position in the designed primer).

Serotypes that could not be distinguished by the presence/absence of genes alone were further analysed for SNP differences, for example serotypes 15B and 15C have the same pseudoDNA sequence therefore SNP differences were detected in their CPS. Clustal Omega Alignment (accessed https://www.ebi.ac.uk/Tools/msa/clustalo/) of the whole CPS cassette revealed SNP differences between these serotypes. Allele-specific primers were designed using PrimerExpress v2.0 software, manually restricted to contain the SNP at the 3' end of the forward or reverse primer. A sub-terminal mismatch primer was designed if the 3' end SNP was either going to be a difference between A or T because real-time PCR may struggle to detect the difference between these two weaker base-pairs (Robertson *et al.,* 2004). For example the true primer sequence reads 5'AACTGTATCA3' while the sub-terminal primer sequence reads 5'AACTGTACCA3' where the thymine (T) has been substituted with a cytosine (C) so

there is a more noticeable difference when the primer binds to a different SNP at the 3' end.

6.2.2 Validation of conventional PCR

Invasive *S. pneumoniae* isolated from children 15 years or younger from January 2007 to December 2012 were resuscitated, isolated and re-cultured (using a single colony) to ensure colonies were genetically identical to each other on the replicate plate. Two ATCC *S. pneumoniae* control strains (ATCC49619 serotype 19F and ATCC6306 serogroup 6) were included in this study. Furthermore, *Streptococcus agalactiae* and *Streptococcus pyogenes* strains were included for testing. Purified DNA was extracted from the two *S. pneumoniae* ATCC control strains, *S. agalactiae* and *S. pyogenes* using QIAamp DNA Mini Kit (50) (Qiagen, Australia).

Minimum SNPs primers (Sigma Life Sciences, Australia) were diluted to 10µM with Ultrapure Distilled water (Invitrogen, USA). Primer sequences are shown in Table 22 of Results. To verify that capsule typing could be applied to the crude DNA thermolysate, conventional PCR was performed using a reaction mix containing PCR Grade Nucleotide Mix (Roche, USA), Taq DNA polymerase (Roche, USA), 10µM forward and reverse primers, $MgCl_2$ (Invitrogen, USA), PCR reaction 10X buffer (Roche, USA) and Ultrapure Distilled water (Invitrogen, USA) with 2µL of the DNA thermolysate. Heat-gradient PCR was performed using a BioRad C1000 Touch Thermal Cycler (BioRad, Australia) as follows: 1min at 95^{o}C, 40 cycles of 95^{o}C for 30sec, gradient temperature 60^{o}C to 50^{o}C for 30sec, and 72^{o}C for 30sec, followed by 10min at 72^{o}C before finally holding at 4^{o}C. To detect the absent or presence of the capsule genes, gel electrophoresis was used (2% agarose, 80V, 400mA, SybrGreen stain).

Later experiments used AmpliTaq Gold® 360 Master Mix (Applied Biosystems, USA), primers and Ultrapure Distilled water for conventional PCR. AmpliTaq Gold® produced clearer bands and reduced the need for optimisation.

Further validation tests were performed to verify if the capsule genes could be amplified in the Queensland *S. pneumoniae* isolates (Table 20). Conventional PCR was performed with the appropriately determined annealing temperature, amplifying all targeted capsule genes from the available Queensland isolates of known serotypes. MinimumSNPs capsule typing was also applied to *Streptococcus pyogenes* and *Streptococcus agalactiae* to determine whether genes could amplify in a different species.

Table 20: *S. pneumoniae* isolates used to validate the initial 23 capsule targets for MinSNPs Capsule Typing.

Serotype has been designated from Quellung reaction; ID (identification) has been assigned to each pneumococcal isolate; Year – year that the pneumococcal isolate was taken from patient suffering from invasive pneumococcal disease (IPD); MLST – sequence type (ST); MLVA4 – MLVA type (MT); Gel # - position of the amplified gene in the gel electrophoresis images contained in the results section of Chapter 6; Notes – mention if the isolate has been associated with the capsule switch.

Serotype	ID	Year	MLST	MLVA4	Gel #	Notes
1	36	2011	306	39	1	Potential capsule switch with serotype 4.
3	8	2011	1230	8	2	
4	157	2009	-	157	3	
6A	224	2008	4237	107	4	Potential capsule switch with serotype 6C.
6B	227	2008	-	227	5	
6C	109	2010	4237	109	6	
7F	21	2011	191	21	7	
8	162	2009	53	162	8	Potential capsule switch with serotype 11A.
9N	164	2009	66	164	9	
9V	307	2007	-	307	10	
10A	282	2007	2068	282	11	
10F	169	2009	4084	169	12	
11A	93	2010	557	93	13	
12F	48	2011	3523	48	14	
14	277	2007	-	277	15	
15B	50	2011	199	50	16	
15C	54	2011	52	53	17	
16F	375	2012	9667	375	18	
18A	55	2011	241	55	19	
18B	291	2007	1255	291	20	Potential capsule switch with serotype 18C.
18C	292	2007	-	292	21	
19A	59	2011	411	59	22	Potential capsule switch with serotypes 15B and 15C.
19F	294	2007	-	294	23	
22A	208	2008	3705	208	24	
22F	143	2009	9663	143	25	
23A	4	2011	42	4	26	
23B	144	2009	9664	144	27	
23F	145	2009	-	145	28	
24F	214	2008	72	214	29	
33B	385	2012	-	385	30	
33F	303	2007	60	217	31	
34	220	2008	9665	220	32	
35B	221	2008	9670	221	33	
35F	152	2009	446	152	34	
38	17	2011	393	17	35	

6.2.3 Real-time PCR and analysis

Real-time PCR (rtPCR) was determined to be a faster and simpler method to type the capsule genes of *S. pneumoniae*, compared to conventional PCR which requires gel electrophoresis for observing the amplified genes. An 18µL reaction containing SybrGreen mastermix (Invitrogen, USA), Ultrapure Distilled water and primers (10µM) was combined with 2µL *S. pneumoniae* DNA thermolysate in real-time PCR tubes. These were placed into a 72-well ring on a real-time PCR Qiagen RotorGene (Qiagen, Australia) or Corbett RotorGene (Corbett, Australia). Real-time PCR conditions were as follows: initialise for 10min at 52^{o}C, cycle 40 times at 95^{o}C for 15sec, 67^{o}C for 20sec and 72^{o}C for 20sec, followed by a standard melt curve (60^{o}C – 90^{o}C) rising 1degree/step, and finally performing a High Resolution Melt curve (60^{o}C – 95^{o}C) rising 0.02degrees/step.

Analysis of rtPCR results included the Melt A. Green heat curve, and the High Resolution Melt (HRM) curve. Since the amplified products are tagged with fluorescent SybrGreen dye, the Melt A. Green setting detects the fluorescent dye and a melt curve is produced when there is an amplified product (present gene). If no product is present (absent gene) then no melt curve should be visible.

However, because rtPCR is known to be highly sensitive, an HRM curve was also performed. During the HRM step, because the melting temperature is raised 0.02degrees/step, this means that we can detect SNP differences within amplified products as they will have slightly different melt curves. To analyse the HRM, a known positive control must be included for each gene in every rtPCR run. The positive control is used to compare the unknown samples and verify if any positive amplifications match the expected positive control – if it does not match, then it is possible that the amplified product is significantly different from the positive control and possibly not a present gene. A normalised difference graph using HRM identifies whether amplified genes are similar to the positive control (within ±5 units), or different.

6.2.4 Double-blind test – rtPCR and conventional PCR

Following the determination of whether rtPCR can be used to amplify capsule genes, a double-blinded test was performed. A sample of 48 *S. pneumoniae* isolates were randomly selected (from 317 isolates) within the restrictions that all Queensland serotypes were represented (35 serotypes) (Table 21). All isolates were re-labelled with a coded number by someone who wasn't directly involved in the experiment. The order in which the isolates were to be tested was randomly assigned by drawing lots

from a container. Each gene was amplified twice per isolate, and all targeted capsule genes were applied. A positive control was included in each run as well as a negative template control and a water control.

All rtPCR results were analysed using the RotorGene Q Series software v2.3.1 (Qiagen, Australia). Stringent criteria had to be met when analysing the absence or presence of the capsule genes; firstly, quantitative analysis was performed by setting the threshold scale on the Cycling A. Green at about half-way on the amplification curves. Any genes that fell below this threshold were considered to be absent. A melt curve that crossed over the threshold was considered to be present. The negative template control and water control always had to fall below this threshold.

Following this, analysis of the Standard melt curve enabled confirmed detection of amplified capsule genes (designated 'present'). The positive control was used for comparison, and any amplified genes that did not match the curve of the positive control were considered to be absent (Figure 40). Curves that appeared to slightly deviate to the left or right of the positive curve were hesitantly considered a present gene because SNPs may be present in these genes which can cause the slight deviations. Sequencing would be required to confirm this.

Finally, analysis of the HRM curve allows even further stringent analysis of those genes that have appeared to be present in the *S. pneumoniae* isolates. Again, the positive control is used for comparison by setting it as the normalised peak in the difference graph, and is observed as a flat line at position '0' (Figure 42). The other genes that deviate more than ±5 units from this normalised line were considered significantly different and therefore labelled an 'absent' gene. However, similar to the standard melt curve, where genes seem to have a nice bell-curve even though it lies outside of the ±5 guidelines, they could be 'present' despite SNP differences. Conventional PCR and gel electrophoresis was also performed on the double-blind to determine whether it can be used for capsule typing.

Table 21: *S. pneumoniae* isolates used for the double-blind validation of the novel capsule typing method.

Double blind code – randomly selected for each isolate by another person; ID (identification) was only revealed after all isolates were capsule typed; serotype is designated based on Quellung serotyping method (performed at QHFSS); Year – year that the pneumococcal isolate was taken from patient suffering from IPD; MLST – sequence type (ST); MLVA4 – MLVA type (MT); Notes – identify whether a possible capsule switch has been previously detected.

Double blind code	ID	Serotype	Year	MLST (ST)	MLVA4 (MT)	Notes
1	290	18A	2007	9082	96	
2	291	18B	2007	1255	291	Potential capsule switch with serotype 18C using MLST.
3	297	22F	2007	1012	209	Potential capsule switch with serotype 33F using MLST.
4	299	23A	2007	42	299	
5	300	23F	2007	-	300	
6	301	23F	2007	-	301	
7	32	7F	2011	191	32	
8	34	9V	2011	-	34	
9	37	1	2011	306	35	
10	47	11A	2011	53	47	Potential capsule switch with serotype 8 using MLST
11	77	19A	2011	-	77	
12	98	18C	2010	1255	98	Potential capsule switch with serotype 18B using MLST.
13	112	6C	2010	4237	107	Potential capsule switch with serotype 6A using MLVA1.
14	167	1	2009	304	167	
15	170	10F	2009	9661	170	
16	172	15B	2009	411	172	Potential capsule switch with serotype 19A and 15C.
17	208	22A	2008	3705	208	
18	385	33B	2012	-	385	
19	386	6A	2012	-	386	
20	214	24F	2008	72	214	
21	231	7F	2008	191	27	
22	233	8	2008	53	233	Potential capsule switch with serotype 11A using MLST.
23	261	12F	2008	3523	48	
24	304	6B	2007	-	304	
25	327	19A	2007	63	60	
26	335	11A	2012	9668	93	
27	264	18C	2008	1255	264	Potential capsule switch with serotype 18B using MLST.
28	279	38	2007	393	279	
29	275	3	2007	9673	11	
30	283	10A	2007	9671	283	
31	284	15B	2007	199	284	Potential capsule switch with serotype 15C using MLST.
32	285	15C	2007	411	59	Potential capsule switch with serotype 15B and 19A using MLVA4.

Double blind code	ID	Serotype	Year	MLST (ST)	MLVA4 (MT)	Notes
33	286	15C	2007	199	286	Potential capsule switch with serotype 15B using MLST.
34	343	19F	2012	-	293	
35	344	23B	2012	439	344	
36	345	33F	2012	60	105	
37	366	33F	2012	1012	366	Potential capsule switch with serotype 22F using MLST.
38	368	35F	2012	1635	368	
39	352	35B	2012	198	402	
40	372	7F	2012	191	22	
41	373	9N	2012	66	373	
42	378	4	2012	306	36	Potential capsule switch with serotype 1.
43	382	14	2012	-	171	
44	374	34	2012	1439	374	
45	375	16F	2012	9667	375	
46	376	22F	2012	698	376	
47	ATCC49619	19F	-	-	-	
48	ATCC6306	6A	-	-	-	

6.3 Results

6.3.1 In silico *analysis of MinimumSNPs*

There was no complete capsule sequence for serotype 6F, 6G, 6H, 11E or 24B available in the NCBI database, therefore only 93 CPS (out of 98) were used for analysis. The mother template consisted of 182 capsule genes that were found associated with 93 serotypes. The 'pseudoDNA sequence' for each of the 93 serotype was manually created for analysis by the MinimumSNPs program (*Appendix A5:* In silico *analysis of 182 targets in the pneumococcal capsule cassette.)* The Minimum SNPs software narrowed this down to 23 capsule genes that would distinguish between most serotypes. Using 23 CPS targets could distinguish 36 out of 93 serotypes, and 39 out of 46 serogroups, with a Simpson's Index of Diversity (D) = 0.9895. The serotypes that formed complexes (i.e. the capsule profiles were exactly the same) were serotypes 7B and 40, serotypes 35A, 35C and 42, serotypes 33A, 33F and 37, and serotypes 12A, 12B, 12C, 44 and 46. A total of 88 primer sets were initially designed to amplify all 23 CPS targets because most gene targets required more than one primer set to ensure all expected serotypes were targeted (Table 22).

The 88 primer sets were then re-analysed using MinimumSNPs to determine whether a smaller set of primers could be used to distinguish the Queensland *S.*

pneumoniae isolates. It was determined that 18 primer sets out of the previous 88 primer sets plus *wzg* internal control (19 primer sets in total) were required to discriminate the 35 Queensland serotypes with a Simpson's Index of Diversity of 0.9933. Not all serotypes were 100% distinguishable including serotypes 22A and 22F, serotypes 19A and 19F, serotypes 18B and 18C, and serotypes 15B and 15C.

Table 22: The 88 primer set identified to amplify 23 capsule genes in *S. pneumoniae*.

Yellow highlighted genes indicates those used for Queensland serotypes only, while all 93 serotypes can be identified using both the yellow highlighted and purple highlighted genes. All primers have been searched in NCBI BLAST to confirm that both forward and reverse primers amplify in expected pneumococcal serotypes. The penalty score, calculated by the primer designing software (a score below 100 is considered ideal for the primers), are all below 20. Accession number – representative isolate taken from NCBI database and used to design primers; Sero. – serotype: L – our designated number for each CPS gene from the 182-gene mother template; Gene – targeted CPS gene; Forward and reverse primer are in 5' to 3' directions; AL – expected length of the amplified product; Tm – melting temperature (^{o}C); Ta – annealing temperature (^{o}C); %gc – guanine (G) and cytosine (C) content in the amplified products; notes – contain details regarded *in silico* amplification of primers in expected serotypes.

Accession	Sero	L	Gene	Forward Primer	Reverse Primer	AL	Tm	Ta	%gc	Notes
CR931639	6B	15	rmlB	TTCGCTTCCACCATGTATCG	GAGGCTGCCTTGGTTGATGA	151	80	58	49	Perfect.
CR931648	9V	65	wcjD	TCAGGCAATTCACCTTCCAAT	GAGCGATCGATTAGAGGATATTATCTC	151	73	54	33	Perfect - only 9A and 9V annealing
CR931681	22A	22	glf	GCCTTCGTTTTGAGCATGAGA	TCATGCGTGATAACTGTCTTTGG	151	76	56	40	Perfect for serotypes 22A, 22F, 17A, 41A, 41F, 16A, 31 and 47A only.
CR931672	18B	22	glf	GCCAGCAGAAGCACAAGCTA	TTTGCCCCACTGTTTCTCTGT	151	75	55	38	Perfect for most serotypes.
CR931706	35C	22	glf	CCGTAGCCTTCGCTTTGATC	CGTGATAACTGTCTTTGCCTGTGT	151	76	56	40	Perfect for serotypes 35C, 42, 33A, 33F and 47F only.
CR931651	10C	22	glf	GAGCAAACGACTCACATGCAA	TGATAAAGGGTGGCAGCTCAGT	151	77	57	44	Perfect for serotype 10C only.
CR931689	25A	22	glf	AATTATCAAGGAAATGCCGTTGTTA	AAGGTTCATCTCCACGTTTCCA	151	74	55	36	Perfect for serotypes 36, 25A, 25F and 38 only.
CR931700	33C	22	glf	TGCCCAGCAAACGGCTAA	AAATGGCGGAAGGTCTGTTG	151	77	57	43	Perfect for serotype 33C only.
CR931633	2	22	glf	GTTGGGCTAGGAGTGGATTGC	CAAGCACTTCGTGCTTAAAACG	151	77	56	42	Perfect for serotype 2 only.
CR931680	21	22	glf	CATTGGAGGGAACATATACTGTGAGA	GTTCGCGACAGGTGAATTGA	151	74	54	35	Perfect for serotypes 21 and 37 only.
CR931666	15F	22	glf	GACGCTCAGCAACGGAACTT	ACTTCAACGTCTTTAAGCATATTTTCG	151	74	55	35	Perfect for serotypes21, 15F, 31, 35C, 42, 33A, 33F, 37, 34, 47F and 39 only.
CR931716	43	22	glf	CATTGGTGGGAACATCTATTGTGA	GTTTGCGACAGGTGAGTTGACA	151	75	56	38	Perfect for serotypes 31, 41A and 43 only.
CR931679	20A	22	glf	CGTCAACTCGCCTGTAGCTAATT	CGGCTCAACATCCTTCATATCA	151	77	56	42	Perfect for serotypes 15F, 20, 31, 33A, 33F, 35A, 35C, 42 and 47F only.
CR931649	10A	22	glf	CATTAAACGCCTTCCAGTTCGT	ACGGTGAGCAAAGAAATCAACA	151	75	55	38	Perfect for serotype 10A and 10B only.
CR931652	10F	22	glf	AGATTGCTGAGCAGACGACTCA	GGTGGCAGCTCAGTAGCAGAA	151	78	57	44	Perfect for serotype 10C and 10F only.
CR931707	35F	22	glf	GCCAGCAGAAGCACAAGTTAAGAT	TTTGCCCCACTGTTTCTCTGT	151	75	55	37	Perfect for serotypes 35F, 29 and degenerative primer for 48, 41F, 40, 35B, 33D, 33B, 24F, 24A, 22A, 22F, 19C, 18F, 18C, 18B, 17A, 17F, 13, 7F, 7A, 7B, 6A, 6B only.
CR931633	2	17	wchA	TTTTTGCTCAGACGCGTATAGG	CGAGGATCATCGTCCACCTT	151	77	56	42	Half of serotypes anneal.
CR931672	18B	17	wchA	AGATGGCGGACCAGCTATTTT	CTTAAACATACCACCTTGCATCGT	151	76	56	40	Most other half resolved.

Accession	Sero	L	Gene	Forward Primer	Reverse Primer	AL	Tm	Ta	%gc	Notes
CR931680	21	17	wchA	GGTGCATTGGTAGGGCTGAT	CACACA TGGAGCGGAACTTG	151	78	57	45	Perfect for serotypes 21 and 8 only.
HM171374	6D	17	wchA	GCAGGTCAGTGGACGAAGTGA	CCCCTATTTGGCTCCATCCT	151	74	55	36	Perfect for serotype 6D and 6C only.
CR931709	37	17	wchA	GGCAGGTCAGTGAACGAAGTG	ATCTACTTCGCTCCATCCTTCATC	151	74	54	34	Perfect for serotype 37 only.
CR931679	20A	17	wchA	CACGCGTCGAAAGAGTAATGG	TCATGGGTAGCGTAGGAAATCA	151	76	56	41	Perfect for serotypes 20A and 20B only.
CR931657	11F	17	wchA	TCGCTCCAT GTGTGTAGATGCT	CTGTGGCAGCTCGTCCAA	151	78	57	44	Perfect for serotypes 11F and 8 only.
CR931654	11B	17	wchA	CATGTGATTGCTAAGCGGGTTA	GGCGGCCATTTTTTCCTATAC	151	78	57	44	Perfect for serotypes 11B, 11C and 23B only.
CR931633	2	12	ugd	GTTTGGATCCCCGTATTGGTT	CGTGTCGAATTGCTCCGTACT	151	78	57	44	Perfect for serotype 2 only.
CR931632	1	12	ugd	AATATCTCGCGGAAAAAGAATTGA	TCAATGACAGCCTCGACTGAA	151	74	55	36	Perfect for serotype 1 only.
CR931713	41A	12	ugd	TAGCCTTACGTGTTTCTTATTTCAATGA	AGGCAATATCCCCCATATCCA	151	75	55	38	Perfect for serotype 41A, 41F, 31 only.
CR931669	17A	12	ugd	TTGGCCTTACGAGTTTCGTATTTC	TTGGCCTTACGAGTTTCGTATTTC	151	75	55	38	Perfect for serotype 17A only.
CR931681	22A	12	ugd	GAATCGGTTATTCAACAGGTTCAGA	CTCTTGATTCGCGCAAAAATT	151	74	54	35	Perfect for serotype 22A, 22F only.
CR931689	25A	12	ugd	CATCTGCCGTAGAACAAGTAATCG	CGAAGAAATTCTGGGCTAAACAG	151	73	54	34	Perfect for 25A, 25F, 38 only.
CR931645	9A	12	ugd	ACTTACCTTGCGTTGCGTGTT	ACCCACCGTAGCCAAAGGAT	151	76	56	41	Perfect for serotypes 9A, 9V, 9N and 9L only.
CR931634	3	12	ugd	GAAAATTGCCATTGCAGGAAGT	CTCAATCGCCTCATCCTTAATTG	151	76	55	39	Perfect for serotype 3 only.
CR931644	8	12	ugd	TGAAGCTGAGGCGGTCAAG	GTGAGAGCCGATTCGTGGAT	151	77	56	42	Perfect for serotype 8 only.
CR931637	5	12	ugd	GTTTGCAACTCTTTTGCAGGAA	CTCCGCATAGGTATCCAATTCG	151	77	55	42	Perfect for serotype 5 only.
CR931633	2	18	wchF	AGGACGAGGCTCCCATTTTT	CGAACTGGCAGGCTCCATT	151	78	58	45	Perfect for serotype 2 only.
CR931671	18A	18	wchF	GATAAGGACCCACGAGTAAAATTTG	TCAGTTTAGTTGATGCCAGAGCTT	151	77	56	42	Perfect for serotypes 24A, 24B, 24F, 48, 18A, 23F, 28A, 28F.
CR931672	18B	18	wchF	CGCTCCACATTAAATAGCAGTGA	TAGAGCTAGACGCCAAGAAACCA	151	75	55	38	Perfect for serotype 16F, 18B, 18C only.
CR931681	22A	18	wchF	AGGACGAGGCTCCCATTTTT	CGAACTGGCAAGCTCCATTT	151	78	57	45	Perfect for serotypes 22A, 22F, 21 only, weak for 2, 32A, 32F.
CR931641	7B	18	wchF	GGGACGATTTGTCCCTGAGA	TGGGTCTTGGTCAAAACCTGTT	151	75	55	37	Perfect for serotypes 7B, 7C, 40 only.
CR931691	27	18	wchF	TTGTGGGAACAGTCTATGACCAA	AAGCCGACATCCAACAACAAG	151	77	56	43	Perfect for serotypes 21, 27 only.
CR931640	7A	18	wchF	GCTTGTCGCATTGGTCCTTT	TGAGCTGCTCTGAAAATTTCCA	151	77	57	44	Perfect for serotypes 7A, 7F only.
CR931683	23A	18	wchF	TGGTGGAACAAATCCTTCGTT	TTGTTCACTCGCTTCAATAACCTT	151	75	55	39	Perfect for serotype 23A only.
CR931684	23B	18	wchF	TGCCGGTTCGAAAGTATTGG	TCCGTCCCGTAAGCAATATAAGTC	151	74	55	35	Perfect for serotype 23B only.
CR931670	17F	18	wchF	ACGGACATGAAGTTGGTGGAA	TCGATTACTTGGGAAAGCTCATC	151	76	56	41	Perfect for 16F, 17F, 18B, 18C only.
CR931674	18F	18	wchF	GGACTGATACAACACGCTCCATAT	AAAAACCCACGTATCATAGCCTCAT	151	74	56	36	Perfect for serotypes 18A, 18F, 28A, 28F only.
CR931696	32A	18	wchF	AGGACGAGGCTCCCATTTTT	CGAACTGGCATGCTCCATTT	151	78	58	46	Perfect for 32A, 32F, degenerative primers 27, 22A, 22F, 21, 2 only.
CR931666	15F	66	wcjE	GCTTGGACTGTATCAATCGGAAA	TTGCCAGAACTCCTTCAAATCC	151	74	55	36	Perfect for some, but others fail to anneal, investigate further.
CR931645	9A	66	wcjE	GTTTTTTGCCCTGTCAGGATACTT	AGAGGCACCAGCGACCTTTT	151	73	54	34	Perfect for serotype 9A, 9V only.
CR931646	9L	66	wcjE	CTGTGTATTAGTTGCTTTTGCGATTT	TTTCAACATTAGAATCCGAATCATG	151	72	53	31	Perfect for all serotypes, weak for 43.
CR931716	43	66	wcjE	AAAGGGTTCGCAATATTTCTAGTAGTCA	ATCCTGATAAAGCAAAGAATACTGGAA	151	72	53	31	Perfect for serotype 43 only.
CR931674	18F	82	gct	TGATTGTAGTTGTTTCAAGTGATGAGTT	AATACGGTACTCTTTAACATCTGACTTTTTC	170	71	53	29	Perfect for serogroup 18, but other serotypes fail to anneal, investigate further.
CR931653	11A	82	gct	AGGAGATTATCTCATTGTTGTAGTTTCAA	CTGTTCCCAGCTTGTTTCAGAA	151	72	53	30	Perfect for serogroup 11.
CR931667	16A	82	gct	TCGTAGAGGCTGTTAGGTATGTTGA	TCAACTCCTTCCTTCTCTAAAAATCA	151	75	55	37	Perfect for serotype 16A only.
CR931718	45	82	gct	TGCTGATATCTTTGTTATTGGTGATG	TTATTTTCCACCTTTGGTCTTGTTT	151	73	53	32	Perfect for serotype 45 only.
CR931641	7B	53	wcwK	GATATTGGCAATTGTCTAGCGGTAA	CTGCATCAAAATTATCTATACTTTCACCAT	165	70	52	26	Perfect for serotype 7B, 7C, 40, but fail to anneal to others.
CR931679	20A	53	wcwK	GATGGTAGGCCGACTTTAAATGA	GATTTGCATCCAGCCATTCA	151	74	55	36	Perfect for serotype 20A, 20B only.
CR931680	21	53	wcwK	CGACCAGCAAAGACCAAGCT	GTGCTCAGAGAGACCAGGAATTC	151	76	56	41	Perfect for 21 only.
CR931704	35A	74	wcrH	TTTTGGATTGGTATGGGATGGT	GAAACTAGCTATTGCTGCCTCTGA	151	76	56	41	Perfect for serotype 35A, 35C, 42, but fail to anneal to others.
CR931705	35B	74	wcrH	GTACGGAGTGCGTGTTGCAA	GTCTGCAATAGCATTTGTTGTTTCA	151	71	53	29	Perfect for serotype 35B only.
CR931694	29	74	wcrH	CCTTTTTGGTCCGAATTTTAATATTG	AGTGTTTCCGCCACAGGTATCT	151	74	54	35	Perfect for serotype 29 only.
CR931708	36	74	wcrH	GTTGCCTACAAGACCGGCATA	ACTACCACCGTCCCAAACCA	151	76	56	41	Perfect for serotype 36 only.
CR931651	10C	74	wcrH	GGGCTTTCCTCTGGT	GGCATTAGCCGACAGTT	151	79	58	47	Perfect for serotypes 10C, 10F

Accession	Sero	L	Gene	Forward Primer	Reverse Primer	AL	Tm	Ta	%gc	Notes
				GGTTT	CCA					only.
CR931716	43	74	wcrH	GGACCTAATTTTACTTCATCGACTAACA	TCCCCAGTAATCCCTGAACAAG	152	73	54	34	Perfect for serotype 43 only.
CR931641	7B	55	rbsF	TGAATAAGACGGGAATACTGTTTACTGA	GGTACCCCAACACCATTCATAAAT	151	72	53	30	Perfect for serotype 7B, 7C, 40, 24F, but fail to anneal to others.
CR931676	19B	55	rbsF	CATAGGAATTGTCATCCTGAGCAA	GATAACACATTTCTTCAGTGCCTAACTT	151	73	54	33	Perfect for 19B, 19C only.
CR931666	15F	98	wchX	AGATACATATTACGAATTGCCGGATT	ACTAGCGGTTCCATTAAAGATTGG	151	71	52	27	Perfect.
CR931700	33C	39	wciN	GGCGTGAAGAACGAATCGAA	CCTCAAAAATGGTCATATAATTGTATCTTG	153	72	53	30	Perfect for 33B, 33C but no others, investigate further.
EF538714	6C	39	wciN	GGCGATGGCGAATTGC	CAAAGCCCTCCCATTTCGA	151	78	57	45	Weakly anneal to 6C, 6D, but no others.
CR931638	6A	39	wciN	AATTTTGTAGATGTATTGAGTGCTTCCA	TTTCTCTCTGCGCAAATTGTTTT	151	71	52	27	Perfect for 6A, 6B, 33D but no others.
CR931681	22A	44	wcwA	CTGGTGAAGAGTGGCGAGTTG	GGATCGGGATTGGTACTTGGT	151	75	55	37	Perfect for all serotypes, however weak anneal for 7A, 7F, 21.
CR931640	7A	44	wcwA	CAAGCACAGTGCGTGAACAA	GCCAGAAAAGCGATAGAATTGG	151	79	58	48	Perfect for 7A, 7F only.
CR931704	35A	73	wciG	TGGTTTTTCGGTGCACTCATAC	TGGTTTTTCGGTGCACTCATAC	151	71	52	28	Perfect for some, weak anneal for some, fail to anneal for some.
CR931667	16A	73	wciG	GGCTTTTCTATTATCTTTTAGGTGGTTT	ATGGTATGCAGTCTTCGCTATGAA	152	71	52	28	Perfect for 16A only, but not others.
CR931707	35F	73	wciG	TTGGTTTTTTGGTGCGCTTAT	TGGCATTTGAAGCACAATATTTG	151	71	52	27	Perfect for 35B, 35F, 47F only.
CR931699	33B	73	wciG	TTGGTTTTTCGGTGCACTAATACTTA	TGGCATCTGAAGTAGGATATTTAATAACTC	151	71	52	28	Perfect for 35A, 35C, 42, 33A, 33F, 20, 33B, 33D, 13
CR931709	37	73	wciG	TTGCATGTGTTGGAGTCGTTCT	CTCTTCCCCAACAATAAATAACCATT	151	74	55	36	Perfect for serotypes 37, 33A and 33F only.
CR931651	10C	73	wciG	GGTGGCAATTGGGTTTATAGTTG	TTTGAGGTTGTCTACATTGAATTGTG	151	73	53	32	Perfect for serotypes 10C and 10F only.
CR931674	18F	105	wcxM	GGTTGTACGTGGAATCGGATTT	GATTCCTTCTTTTGACAGTGCATCT	151	74	55	36	Perfect.
CR931669	17A	112	wcrT	GACATATTCAGCCTGAAGTTCCATT	TCCAGAAGGAGCTTTCCATCA	151	73	54	33	Perfect for 17A only.
CR931670	17F	112	wcrT	GGCAGAATTGGCTAAAAAGTATGG	TTCATTTGCCCAACTCAAACAA	151	72	53	30	Perfect for 17F only.
CR931700	33C	67	wcjG	GGGATGTGGCAAGTTTCTGGTA	AATGAGCCCCTTCTCCAGACA	151	76	56	40	Perfect.
CR931700	33C	67	wcjG	TATTGGACCTAGACCGATTTTGG	GCTCCATTTCACAGCGTTCA	152	77	57	43	Perfect.
CR931633	2	9	wzy	TGGAACTCTTAGAAATTGGGAAAATT	AAGGAAACTTGTCACTGTTGAATCAG	151	73	54	33	Serotype specific, investigate further.
CR931635	4	29	wciK	CGGGAATCACACTCCCTTACTAGA	ATCCTTCATAGGTAAAACGAAACCA	151	73	54	34	Perfect for serotype 4.
CR931688	24F	140	wcxI	AATGATGTGCCTGATACTTTTATTATGC	TTTGCTTAGCTGACACCATTCC	151	72	53	31	Perfect for serogroup 24.
CR931689	25A	149	wcyE	TTGAGGTGCGATTTTTATATTGTGAA	TGTTTCCATCTTGCTTCTCCAGTA	151	72	53	30	Perfect.
CR931691	27	152	whaK	AGCAGTATCTTTTGGATGGCAAGT	AAGGTCCCATCGCTGACAAT	151	76	56	41	Perfect.
CR931688	24F	CON	wzg	ATTAGTTCGGTGTCGCGATCA	TGCCCGCATGGGTTAATTTA	151	77	57	43	Perfect.

Using the 19 primer set (18 primers plus *wzg*) for Queensland isolates, a barcode sequence (only contains the targeted genes) was devised for each serotype previously detected in Queensland (Table 23). These barcodes were the expected capsule profiles for the Queensland invasive *S. pneumoniae*, where T represents a present gene and A represents an absent gene. A number of serotypes appear to have the same barcode or capsule profile, including serotypes 22A and 22F, serotypes 19A and 19F, serotypes 18B and 18C, and serotypes 15B and 15C. These serotypes can only be differentiated using SNPs. Research into the NCBI database revealed several SNP differences within the capsule genes from which SNP-specific primers were designed (Table 24 and 25). These SNP-specific primers can be used to differentiate between those serotypes with the same capsule profile.

Table 23: MinimumSNPs 19 primer set for serotyping the 35 *S. pneumoniae* serotypes found in Queensland, and the barcode sequences for Queensland serotypes using the 19 primer set.

The top table shows the 19 primer sets for the Queensland isolates. Serotype of NCBI pneumococcal isolate; Locus – our designated number for each CPS gene from the 182-gene mother template; Gene – targeted CPS gene; Forward and reverse primer are in 5' to 3' directions; Reverse complement of the reverse primer is given. The bottom table displays the 'barcode' CPS profile for each of the 35 serotypes previously detected in Queensland. The 19 primers are listed across the top, and the *in silico* results of expected presence (highlighted pink 'T') of a gene, or absence (not highlighted 'A') of a gene.

Serotype	Locus	Gene	Forward Primer	Reverse Primer	Reverse primer (reverse complement)
6B	15	rmlB	GACAATCTTGACAAAAGGGCAAA	GTCATGTCCTGCACGGTCAGT	ACTGACCGTGCAGGACATGAC
25A	22	glf	AATTATCAAGGAAATGCCGTTGTTA	AAGGTTCATCTCCACGTTTCCA	TGGAAACGTGGAGATGAACCTT
10A	22	glf	CATTAAACGCCTTCCAGTTCGT	ACGGTGAGCAAAGAAATCAACA	TGTTGATTTCTTTGCTCACCGT
35F	22	glf	GCCAGCAGAAGCACAAGTTAAGAT	TTTGCCCCACTGTTTCTCTGT	ACAGAGAAACAGTGGGGCAAA
2	17	wchA	TTTTTGCTCAGACGCGTATAGG	CGAGGATCATCGTCCACCTT	AAGGTGGACGATGATCCTCG
18B	17	wchA	AGATGGCGGACCAGCTATTTT	CTTAAACATACCACCTTGCATCGT	ACGATGCAAGGTGGTATGTTTAAG
21	17	wchA	GGTGCATTGGTAGGGCTGAT	CACACATGGAGCGGAACTTG	CAAGTTCCGCTCCATGTGTG
6D	17	wchA	GCAGGTCAGTGGACGAAGTGA	CCCCTATTTGGCTCCATCCT	AGGATGGAGCCAAATAGGGG
9A	12	ugd	ACTTACCTTGCGTTGCGTGTT	ACCCACCGTAGCCAAAGGAT	ATCCTTTGGCTACGGTGGGT
3	12	ugd	GAAAATTGCCATTGCAGGAAGT	CTCAATCGCCTCATCCTTAATTG	CAATTAAGGATGAGGCGATTGAG
18A	18	wchF	GATAAGGACCCACGAGTAAAATTTG	TCAGTTTAGTTGATGCCAGAGCTT	AAGCTCTGGCATCAACTAAACTGA
18B	18	wchF	CGCTCCACATTAAATAGCAGTGA	TAGAGCTAGACGCCAAGAAACCA	TGGTTTCTTGGCGTCTAGCTCTA
9L	66	wcjE	CTGTGTATTAGTTGCTTTTGCGATTT	TTTCAACATTAGAATCCGAATCATG	CATGATTCGGATTCTAATGTTGAAA
15F	98	wchX	AGATACATATTACGAATTGCCGGATT	ACTAGCGGTTCCATTAAAGATTGG	CCAATCTTTAATGGAACCGCTAGT
6A	39	wciN	AATTTTGTAGATGTATTGAGTGCTTCCA	TTTCTCTCTGCGCAAATTGTTTT	AAAACAATTTGCGCAGAGAGAAA
33B	73	wciG	TTGGTTTTTCGGTGCACTAATACTTA	TGGCATCTGAAGTAGGATATTTAATAACTC	GAGTTATTAAATATCCTACTTCAGATGCCA
33C	67	wcjG	GGGATGTGGCAAGTTTCTGGTA	AATGAGCCCCTTCTCCAGACA	TGTCTGGAGAAGGGGCTCATT
4	29	wciK	CGGGAATCACACTCCCTTACTAGA	ATCCTTCATAGGTAAAACGAAACCA	TGGTTTCGTTTTACCTATGAAGGAT
24F	CON	wzg	ATTAGTTCGGTGTCGCGATCA	TGCCCGCATGGGTTAATTTA	TAAATTAACCCATGCGGGCA

Serotype	rmlB L15	wchA L17-2	glf L22-35F	wcjE L66-9L	wchA L17-18B	wciG L67-33C	wchF L18-18A	glf L22-25A	wchF L18-18B	wciG L73-33B	wchX L98	glf L22-10A	wchA L17-21	wchA L17-6D	ugd L12-9A	ugd L12-3	wciN L39-6A	wciK L29	wzg
6B	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T
22A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
23F	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
23B	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
6C	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
6A	T	T	T	A	A	A	A	A	A	A	A	A	A	T	A	A	T	A	T
19A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
19F	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
24F	T	A	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	T
18B	T	A	T	A	T	A	T	A	T	A	A	A	A	A	T	A	A	A	T
18C	T	A	T	A	T	A	T	A	T	A	A	A	A	A	T	A	A	A	T
7F	T	A	T	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	T
19A	T	A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	T
16F	T	A	A	A	T	A	A	A	T	A	T	A	A	A	A	A	A	A	T
23A	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	T
1	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
33B	A	A	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	T
35B	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
9V	A	A	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	T
31A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	T
33F	A	A	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	T
34	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
35F	A	A	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	T
9N	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	T
15B	A	A	A	A	T	A	T	A	A	A	T	A	A	A	A	A	A	A	T
15C	A	A	A	A	T	A	T	A	A	A	T	A	A	A	A	A	A	A	T
14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
10A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T
10F	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	T
38	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T
8	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	T	A	A	T
3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
12F	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T

Table 24: Identification of SNP differences within *S. pneumoniae* capsule genes.

Accession number was obtained from the NCBI database. 'Sequence' numbers indicate the location of the SNP and the 'amino acid' indicates whether the SNP difference changes the amino acid produced when translated.

Serotype	CPS gene	Accession #	SNP	Sequence	Amino Acid
22A	*wcwA*	CR931681	1bp	8521-8536	Alanine
22F		CR931682	difference	8521-8540	Threonine
19A	*wchA*	CR931675	Lots of SNPs		
19F		CR931678			
18B	*wzd*	CR931672	1bp	3398-3412	Threonine
18C		CR931673	difference	3395-3413	Alanine
15B	*wzh*	CR931664	4bp	2922-2938	Lysine
15C		CR931665	differences	2924-2937	Arginine
15B	*wchM*	CR931664		10189-10200	
15C		CR931665		10187-10200	
15B	*Unknown*	CR931664		12936-12955	
15C		CR931665		12938-12954	
15B	*gtp3*	CR931664		17414-17432	Valine
15C		CR931665		17417-17434	Isoleucine
28A	*wchF*	CR931692	2bp	6897-6909	Alanine
28F		CR931693	difference	6586-6604	threonine
28A	*wchF*	CR931692		7679-7693	Leucine
28F		CR931693		7368-7386	Isoleucine
32A	*tnp*	CR931696	1bp	22678-22695	
32F		CR931697	difference	22774-22789	
7A	*wzg*	CR931640	Lots of SNPs;	1551-1565	Arginine
7F		CR931643	5 SNPs	1579-1597	Glycine
7A	*wzd*	CR931640	examined	3910-3924	
7F		CR931643		3939-3954	
7A	*tnp*	CR931640		23569-23605	
7F		CR931643		23573-23609	
7A	*tnp*	CR931640		23612-23641	
7F		CR931643		23616-23645	
7A	*tnp*	CR931640		23731-23784	
7F		CR931643		23735-23787	
20A	*wzg*	CR931679	4bp	1529-1542	
20B		JQ653093	difference	329-342	
20A	*wzg*	CR931679		1586-1597	Threonine
20B		JQ653093		387-400	Isoleucine
20A	*wzg*	CR931679		1768-1781	Proline
20B		JQ653093		570-582	leucine
20A	*wcjE*	CR931679		17982-17995	
20B		JQ653093		16784-16797	
10A	*wcjG*	CR931649	1bp	5490-5500	
10B		CR931650	difference	5194-5211	
7B	*glf*	CR931641	13bp		
40		CR931712	differences		
9A	*wze*	CR931645	3bp	4716-4730	
9V		CR931648	differences	5035-5046	
9A	*wcjB*	CR931645		10788-10802	Glutamic

Serotype	CPS gene	Accession #	SNP	Sequence	Amino Acid
9V		CR931648		11104-11118	Lysine
9A	*wcjE*	CR931645		18525-18538	
9V		CR931648		18841-18854	
25A	*glf*	CR931689-	8bp	4882-4891	Threonine
25F		CR931690	differences	4851-4862	Methionine
25A	*wzg*	CR931689-		9042-9054	Arginine
25F		CR931690		9014-9024	Glutamine
25A	*wzg*	CR931689-		9152-9162	
25F		CR931690		9123-9134	
25A	*wzg*	CR931689-		9198-9208	Aspartic
25F		CR931690		9169-9180	Asparagine
25A	*tnp*	CR931689-		27221-27232	
25F		CR931690		27193-27205	
25A	*glf*	CR931689-		4578-4590	Glycine
25F		CR931690		4549-4559	Aspartic
25A	*tnp*	CR931689-		22995-23006	
25F		CR931690		22966-22975	
25A	*tnp*	CR931689-		26562-26573	
25F		CR931690		3888-3899	
33A	*wze*	CR931698	3bp	4093-4102	Serine
33F		CR931702	differences	4093-4104	Proline
	wcjE			16166-16177	
				16165-16179	
	tmp			18258-18267	Leucine
				16837-16846	Phenylalanine

For the serotypes that cannot be differentiated using the 19 primer set, SNP-specific primers have been designed to further discriminate these serotypes (Table 25). Sub-terminal mismatch forward primers were designed, meaning the 3' end for each serotype contains the SNP.

Table 25: SNP-specific primer pairs for *S. pneumoniae* serotypes.

Serotypes	Gene	Forward primer	Reverse Primer	Forward sub-terminal mismatch	Reverse sub-terminal mismatch
22A	wcwA	CGAATTGAATATTATGAGCATACGG	GGTAGTGGATCGGGATTGGTACT	CGAATTGAATATTATGAGCATAAGG	GGTAGTGGATCGGGATTGGTACT
22F	wcwA	CGAATTGAATATTATGAGCATACGA	GGTAGTGGATCGGGATTGGTACT	CGAATTGAATATTATGAGCATAAGA	GGTAGTGGATCGGGATTGGTACT
19A	wchA	CGTATAGGAAAAAATGGCCGCC	TTGCATGGTATTTTGTTCCATGAG	CGTATAGGAAAAAATGGCCACC	TTGCATGGTATTTTGTTCCATGAG
19F	wchA	CGTATAGGAAAAAATGGCCGCT	TTGCATGGTATTTTGTTCCATGAG	CGTATAGGAAAAAATGGCCACT	TTGCATGGTATTTTGTTCCATGAG
18B	wzd	GGAGAAAATATGCAAGATCAAAACA	AGCAACTGAAGAAGTTATAATTGCCACTA	GGAGAAAATATGCAAGATCAAAGCA	AGCAACTGAAGAAGTTATAATTGCCACTA
18C	wzd	GGAGAAAATATGCAAGATCAAAACG	AGCAACTGAAGAAGTTATAATTGCCACTA	GGAGAAAATATGCAAGATCAAAGCG	AGCAACTGAAGAAGTTATAATTGCCACTA
15B	wzh	GTTCTGGATAAGCTGGAAAAAAA	AACTCTATCAAGGCATAACGACTATCAT	GTTCTGGATAAGCTGGAAAAGAA	AACTCTATCAAGGCATAACGACTATCAT
15C	wzh	GTTCTGGATAAGCTGGAAAAAAG	AACTCTATCAAGGCATAACGACTATCAT	GTTCTGGATAAGCTGGAAAAGAG	AACTCTATCAAGGCATAACGACTATCAT
28A	wchF	GAATTCAAGCAATAGGTGGGG	CGAACAGGTAAACTCCATTTTTCTC	GAATTCAAGCAATAGGTGAGG	CGAACAGGTAAACTCCATTTTTCTC
28F	wchF	GAATTCAAGCAATAGGTGGGA	CGAACAGGTAAACTCCATTTTTCTC	GAATTCAAGCAATAGGTGAGA	CGAACAGGTAAACTCCATTTTTCTC
32A	tnp	GGATGAAACTTCCTTATAGGGTGCTA	AGCGTAATCCCTTGTTTCTCTGA	GGATGAAACTTCTTATAGGGTGTTA	AGCGTAATCCCTTGTTTCTCTGA
32F	tnp	GGATGAAACTTCTTATAGGGTGCTG	AGCGTAATCCCTTGTTTCTCTGA	GGATGAAACTTCTTATAGGGTGTTG	AGCGTAATCCCTTGTTTCTCTGA
7A	wzd	CTCGGATTTATGTAGTTAACCGTGG	GTATGATCCTGCCTGCAAGTCTT	CTCGGATTTATGTAGTTAACCGGGG	GTATGATCCTGCCTGCAAGTCTT
7F	wzd	CTCGGATTTATGTAGTTAACCGTGA	GTATGATCCTGCCTGCAAGTCTT	CTCGGATTTATGTAGTTAACCGGGA	GTATGATCCTGCCTGCAAGTCTT
20A	wzg	CGTATGATGTTCAAGGTATAGGTGTTAAC	AAACGGTATTAACACTTCGCTTCA	CGTATGATGTTCAAGGTATAGGTGTTGAC	AAACGGTATTAACACTTCGCTTCA
20B	wzg	CGTATGATGTTCAAGGTATAGGTGTTAAT	AAACGGTATTAACACTTCGCTTCA	CGTATGATGTTCAAGGTATAGGTGTTGAT	AAACGGTATTAACACTTCGCTTCA
10A	wcjG	CCTCGTCTTACTAATATCGGAGATTTTATA	CAATAAGGCTCATATCACCTAACATC	CCTCGTCTTACTAATATCGGAGATTTTATA	CAATAAGGCTCATATCACCTAACGTC
10B	wcjG	CCTCGTCTTACTAATATCGGAGATTTTATT	CAATAAGGCTCATATCACCTAACATT	CCTCGTCTTACTAATATCGGAGATTTTATA	CAATAAGGCTCATATCACCTAACGTT
7B	glf	GAAGAACAGGCGATTTCTCTTGTAGG	TGCCCCACTGTTTCTCTGTATAGTC	GAAGAACAGGCGATTTCTCTTGTGGG	TGCCCCACTGTTTCTCTGTATAGTC
40	glf	GAAGAACAGGCGATTTCTCTTGTAGA	TGCCCCACTGTTTCTCTGTATAGTC	GAAGAACAGGCGATTTCTCTTGTGGA	TGCCCCACTGTTTCTCTGTATAGTC
9A	wze	CGCTTTGTGCACGAACCTACA	TTTCCTTCTCCTAGTTTCACAGAAGTG	CGCTTTGTGCACGAACCTGCA	TTTCCTTCTCCTAGTTTCACAGAAGTG
9V	wze	CGCTTTGTGCACGAACCTACG	TTTCCTTCTCCTAGTTTCACAGAAGTG	CGCTTTGTGCACGAACCTGCG	TTTCCTTCTCCTAGTTTCACAGAAGTG
25A	glf	CCTTTCAATATGAATACTTTCTATGCTAC	TTCGGTTCAACGTCTTTCATGT	CCTTTCAATATGAATACTTTCTATGCGAC	TTCGGTTCAACGTCTTTCATGT
25F	glf	CCTTTCAATATGAATACTTTCTATGCTAT	TTCGGTTCAACGTCTTTCATGT	CCTTTCAATATGAATACTTTCTATGCGAT	TTCGGTTCAACGTCTTTCATGT
11A	wcjE	CACTAAATACGGGAAAGATAGTTTAGGT	CAACATTAGAATCCGAATCATGCT	CACTAAATACGGGAAAGATAGTTTAAGT	CAACATTAGAATCCGAATCATGCT
11D	wcjE	CACTAAATACGGGAAAGATAGTTTAGGG	CAACATTAGAATCCGAATCATGCT	CACTAAATACGGGAAAGATAGTTTAAGG	CAACATTAGAATCCGAATCATGCT
33A	wze	GGAGATATTCGCAATTCTGTTATGT	GTCTGTAGTTCCTGATAAAAAATTCTGTCA	GGAGATATTCGCAATTCTGTTAGGT	GTCTGTAGTTCCTGATAAAAAATTCTGTCA
33F	wze	GGAGATATTCGCAATTCTGTTATGC	GTCTGTAGTTCCTGATAAAAAATTCTGTCA	GGAGATATTCGCAATTCTGTTAGGC	GTCTGTAGTTCCTGATAAAAAATTCTGTCA
11B	wchA	GTCATAGCCCAGAGTTTTCTTGTC	TTGAAACAATCTCTGTTTCTCTCACA	GTCATAGCCCAGAGTTTTCTTATC	TTGAAACAATCTCTGTTTCTCTCACA
11C	wchA	GTCATAGCCCAGAGTTTTCTTGTT	TTGAAACAATCTCTGTTTCTCTCACA	GTCATAGCCCAGAGTTTTCTTATT	TTGAAACAATCTCTGTTTCTCTCACA
9L	wcjE	GCTTGGTTCTTGTTTTATGAAGTAGA	GCAACTAATACACAGACAATAAAAGCAA	GCTTGGTTCTTGTTTTATGAAGTGGA	GCAACTAATACACAGACAATAAAAGCAA
9N	wcjE	GCTTGGTTCTTGTTTTATGAAGTAGG	GCAACTAATACACAGACAATAAAAGCAA	GCTTGGTTCTTGTTTTATGAAGTGGG	GCAACTAATACACAGACAATAAAAGCAA
12A	wcil	GGATGGGCACAAATTAATGGA	GGGATTGTTCTAAGGATACATCTAAGATC	GGATGGGCACAAATTAATAGA	GGGATTGTTCTAAGGATACATCTAAGATC
12B	wcil	GGATGGGCACAAATTAATGGA	GGGATTGTTCTAAGGATACATCTAAGATC	GGATGGGCACAAATTAATAGA	GGGATTGTTCTAAGGATACATCTAAGATC
12F	wcil	GGATGGGCACAAATTAATGGG	GGGATTGTTCTAAGGATACATCTAAGATC	GGATGGGCACAAATTAATAGG	GGGATTGTTCTAAGGATACATCTAAGATC
44	wcil	GGATGGGCACAAATTAATGGG	GGGATTGTTCTAAGGATACATCTAAGATC	GGATGGGCACAAATTAATAGG	GGGATTGTTCTAAGGATACATCTAAGATC
46	wcil	GGATGGGCACAAATTAATGGA	GGGATTGTTCTAAGGATACATCTAAGATC	GGATGGGCACAAATTAATAGA	GGGATTGTTCTAAGGATACATCTAAGATC
12A	wcxE	CATCATGTGAAGATTATTCGCTTG	TAGCCGAAATAAGCTTTCCAGAA	CATCATGTGAAGATTATTCGCGTG	TAGCCGAAATAAGCTTTCCAGAA
12B	wcxE	CATCATGTGAAGATTATTCGCTTG	TAGCCGAAATAAGCTTTCCAGAA	CATCATGTGAAGATTATTCGCGTG	TAGCCGAAATAAGCTTTCCAGAA
12F	wcxE	CATCATGTGAAGATTATTCGCTTG	TAGCCGAAATAAGCTTTCCAGAA	CATCATGTGAAGATTATTCGCGTG	TAGCCGAAATAAGCTTTCCAGAA
44	wcxE	CATCATGTGAAGATTATTCGCTTG	TAGCCGAAATAAGCTTTCCAGAA	CATCATGTGAAGATTATTCGCGTG	TAGCCGAAATAAGCTTTCCAGAA
46	wcxE	CATCATGTGAAGATTATTCGCTTA	TAGCCGAAATAAGCTTTCCAGAA	CATCATGTGAAGATTATTCGCGTA	TAGCCGAAATAAGCTTTCCAGAA
12A	wcxB	GCATTTACAGGACACTCGGCTC	ATGGTCACAAACCCCCTTTG	GCATTTACAGGACACTCGGATC	ATGGTCACAAACCCCCTTTG
12B	wcxB	GCATTTACAGGACACTCGGCTT	ATGGTCACAAACCCCCTTTG	GCATTTACAGGACACTCGGATT	ATGGTCACAAACCCCCTTTG
12F	wcxB	GCATTTACAGGACACTCGGCTT	ATGGTCACAAACCCCCTTTG	GCATTTACAGGACACTCGGATT	ATGGTCACAAACCCCCTTTG
44	wcxB	GCATTTACAGGACACTCGGCTT	ATGGTCACAAACCCCCTTTG	GCATTTACAGGACACTCGGATT	ATGGTCACAAACCCCCTTTG
46	wcxB	GCATTTACAGGACACTCGGCTC	ATGGTCACAAACCCCCTTTG	GCATTTACAGGACACTCGGATC	ATGGTCACAAACCCCCTTTG

6.3.2 Optimisation of capsule typing on conventional PCR

Initial validation of whether crude DNA thermolysate could be used for capsule typing demonstrated that amplification should be no issue. The *wzg* was successfully amplified using a heat-gradient PCR for serotype 7F and 19A, both common serotypes discovered in Queensland (Figure 36). Further validation using heat-gradient identified that the most appropriate annealing temperature for all primers was 67°C (Figure 37).

Figure 36: Heat-gradient PCR of isolate 21SP11 (ID 21) serotype 7F (top) and 57SP11 (ID 57) serotype 19A (bottom) using thermolysate DNA.

Lane 2, DNA marker ladder VIII. The heat-gradient temperatures are indicated in the key.

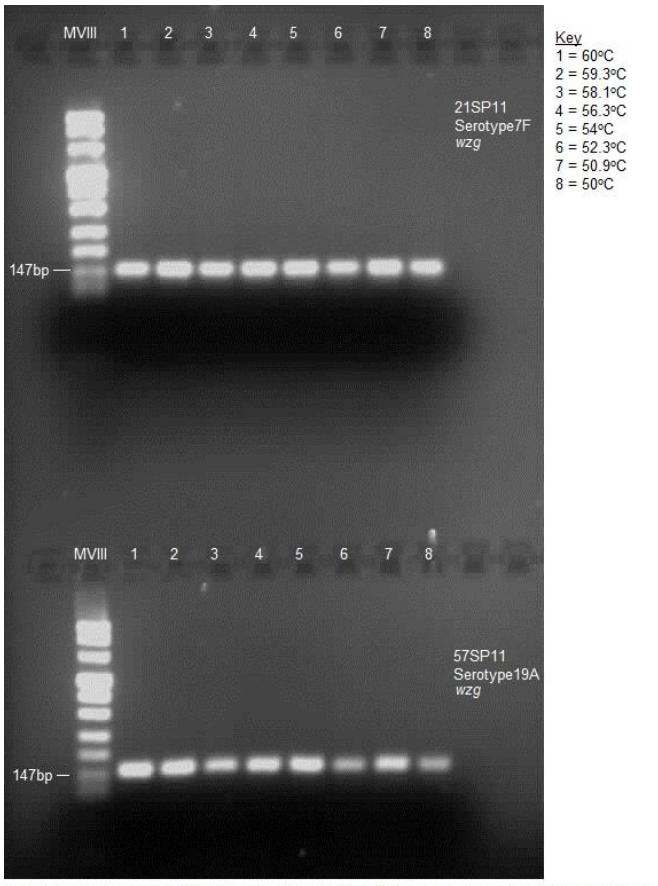
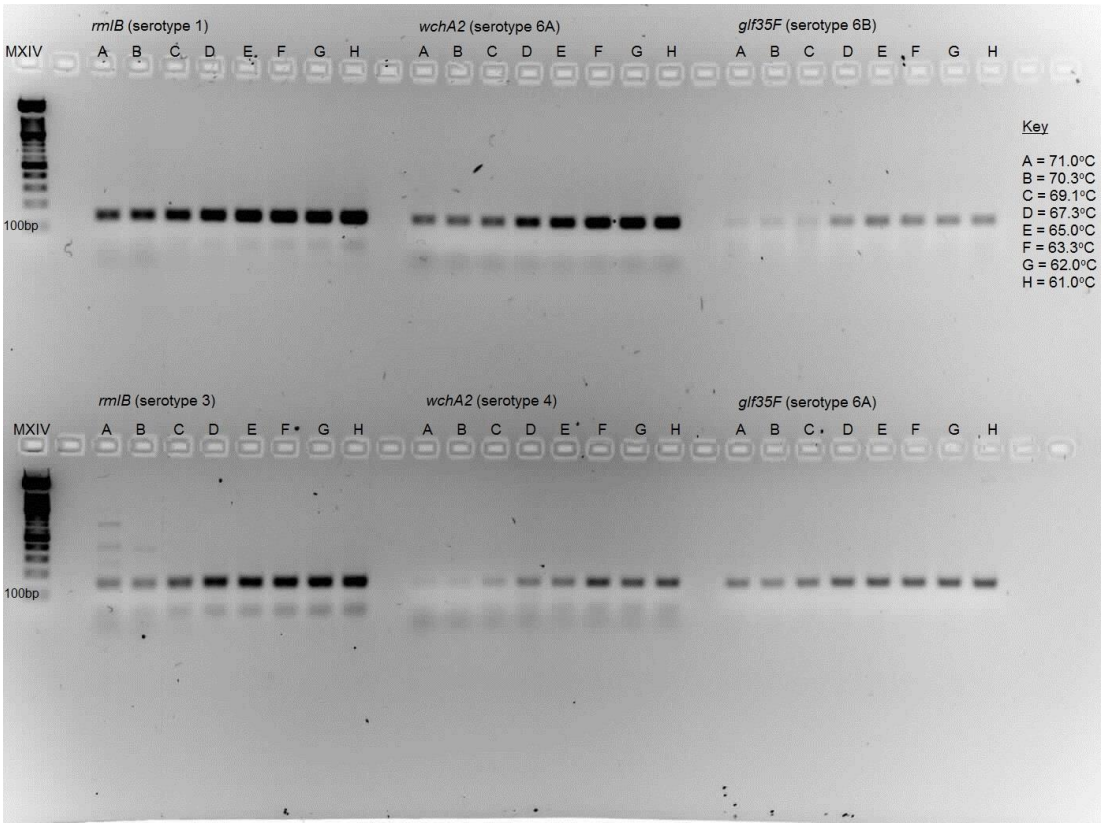


Figure 37: Heat-gradient PCR performed on *S. pneumoniae* serotypes using rmlB, wchA, and glf capsule genes identified in this study.

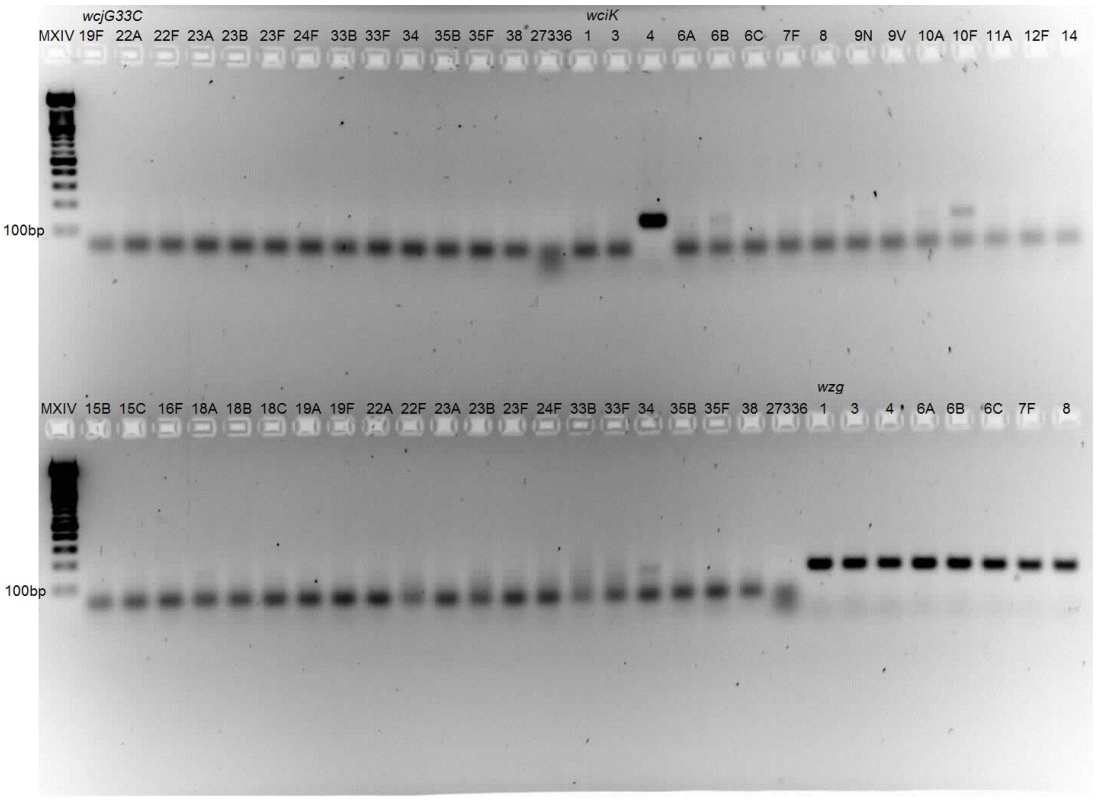
A DNA marker ladder VIII was used in lane 1 (top and bottom). The heat-gradient temperatures are shown in the key. Two *S. pneumoniae* serotypes per gene is shown, where the top row contains positive control (gene should be present in these serotypes), and the bottom row contains serotypes that the gene should be absent.



Following this, amplification of the 19 CPS genes was validated (Table 23) in all Queensland clinical isolates that were available was performed using conventional PCR, this time with the annealing temperature at 67°C. Gel electrophoresis was performed to visualise the presence or absence of the genes. All present bands were expected to be visualised at 151bp in size, based from *in silico* results in Table 22. Smaller and fainter bands were thought to be primer-dimers, and were confirmed to be <100bp as seen on the gels. Where bands were expected to be present, based from the *in silico* analysis previously performed, we could see a present band, for example *wciK* is only expected to amplify in serotype 4, which is what we have observed (Figure 38). An internal control gene *wzg* was also included for identification of *S. pneumoniae*. This validation has confirmed that all *S. pneumoniae* serotypes have a present *wzg* (*Appendix A6: Determination of presence/absence of 19 capsule targets in* S. pneumoniae *using conventional PCR and gel electrophoresis* – in gel 12 and 13).

Figure 38: Validation of capsule genes *wcjG33C*, *wciK* and *wzg* on *S. pneumoniae* Queensland serotypes using gel electrophoresis.

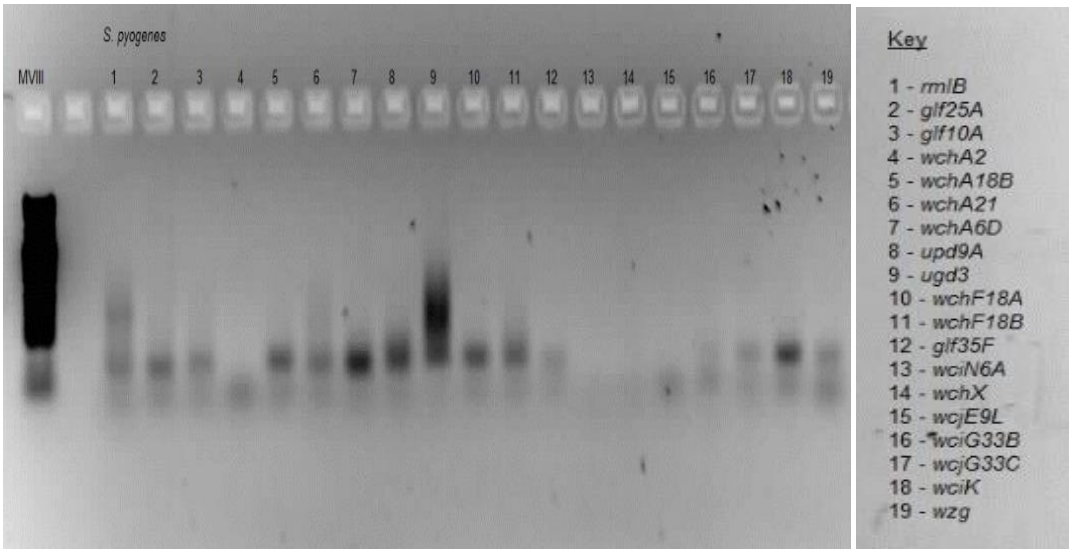
A DNA marker ladder XIV was used in lane 1 (top and bottom). Bands falling below 100bp are primer-dimers. A strong band at 100-200bp indicate a positive amplification.



In comparison, the *wzg* gene was not amplifiable in *Streptococcus pyogenes* (Figure 39) or *S. agalactiae*. Similarly, the other capsular genes could were not amplifiable in *S. pyogenes* or *S. agalactiae*.

Figure 39: MinimumSNPs capsule typing applied to *Streptococcus pyogenes*.

A DNA marker ladder VIII was used in lane 1. Lane 2 is blank. The respective capsule genes amplified are shown in the key.



Generally, where genes were expected to amplify in pneumococcal isolates, a band was visualised. However some CPS bands amplified when they were not expected to amplify. Closer examination revealed that even though when examining the NCBI database several primers pairs for one gene was serotype specific, in reality they could amplify more serotypes. For example, *wchA21* is amplified in several serotypes unexpectedly, however if we examine *wchA2* (different primer pair for the same gene) the *wchA* gene is actually present (Table 26). Therefore we had to accommodate the fact that as long as one of the primer-set identifies a present gene; we could determine that the gene was indeed present. A summary of the amplification of non-amplified Minimum SNPs targets indicated that the majority of expected amplifications were observed, however a number of unexpected amplifications were also observed (Table 26). The raw data is available in *Appendix A6: Determination of presence/absence of 19 capsule targets in* S. pneumoniae *using conventional PCR and gel electrophoresis.* In addition, one of the ATCC strain failed to amplify the *wzg* which should be present in all pneumococci – it is thought that due to it being a laboratory strain that has undergone numerous sub-cultures, it is possible that the *wzg* has been lost from this strain. Sequencing would be required to determine this, which has currently not been performed in this study due to limited funding.

Table 26: Summary of amplification of MinimumSNPs 19 targets in 35 Queensland *S. pneumoniae* serotypes.

T = present gene, and A = absent gene. Green highlighted genes are expected presence of the gene. Dark orange highlight represents unexpected amplification of a gene. Light orange represents unexpected non-amplification of a gene. Allele file code was used for submission of pseudoDNA sequences into MinimumSNPs.

Serotype	allele file code	wzg	rmlB L15	glf L22-25A	wchA L17-21	wchA L17-18B	wchA L17-6D	ugd L12-3	wchA L17-2	wciK L29	glf L22-35F	wciN L39-6A	wcjE L66-9L	ugd L12-9A	glf L22-10A	wcjG L67-33C	wchX L98	wchF L18-18A	wchF L18-18B	wciG L73-33B
19F	ATCC49619	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
6A	ATCC6306	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A
1	cps_>1	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
7F	cps_>13	T	T	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A
8	cps_>14	T	A	A	T	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A
9N	cps_>17	T	T	A	A	T	A	A	T	A	A	A	T	T	A	A	A	A	A	T
9V	cps_>18	T	T	A	T	T	A	A	T	A	A	A	T	T	A	A	A	A	A	A
10A	cps_>19	T	T	T	A	T	A	A	A	A	A	A	T	T	T	T	A	A	A	A
10F	cps_>22	T	T	A	A	T	A	A	A	A	A	A	T	T	T	T	A	A	A	A
11A	cps_>23	T	T	A	A	T	T	A	T	A	A	A	T	T	A	A	A	A	A	A
3	cps_>3	T	T	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A
12F	cps_>31	T	T	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A
14	cps_>33	T	A	A	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A
15B	cps_>35	T	T	A	A	T	A	A	T	A	A	A	A	A	A	A	T	A	A	A
15C	cps_>36	T	T	A	A	T	A	A	T	A	A	A	A	A	A	A	T	A	A	A
16F	cps_>39	T	T	A	T	T	T	A	T	A	T	A	T	T	A	A	A	T	T	A
4	cps_>4	T	T	A	A	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A
18A	cps_>42	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A
18B	cps_>43	T	T	A	A	T	T	A	T	A	T	A	T	A	T	A	T	T	T	A
18C	cps_>44	T	T	A	A	T	A	A	T	A	T	A	T	A	A	A	A	A	T	A
19A	cps_>46	T	T	A	A	A	T	A	T	A	T	A	A	A	A	A	A	A	T	A
19F	cps_>49	T	T	A	T	T	A	A	T	A	T	A	A	A	T	A	A	A	T	A
22A	cps_>53	T	T	T	T	T	A	A	T	A	T	A	A	A	A	A	A	T	T	A
22F	cps_>54	T	T	A	T	A	T	A	T	A	T	A	A	A	A	A	T	A	T	A
23A	cps_>55	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	T	A	A	A
23B	cps_>56	T	T	A	T	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A
23F	cps_>57	T	T	A	T	T	A	A	T	A	A	A	A	A	T	A	T	T	T	A
6A	cps_>6	T	T	A	T	T	T	A	T	A	T	A	A	A	A	A	A	A	T	A
24F	cps_>60	T	T	A	A	T	A	A	T	A	T	A	A	A	T	A	A	T	A	A
6B	cps_>7	T	T	A	T	T	A	A	T	A	T	T	A	A	A	A	A	A	A	A
33B	cps_>71	T	T	A	T	A	A	A	T	A	T	A	A	A	A	A	A	T	A	T
33F	cps_>74	T	T	T	T	A	T	A	T	A	T	A	T	T	T	A	A	T	A	A
34	cps_>75	T	T	T	T	T	A	A	T	A	A	A	A	T	T	A	A	T	A	A
35B	cps_>77	T	T	T	T	A	T	A	T	A	T	A	A	A	T	A	A	A	A	A
35F	cps_>79	T	T	A	A	A	A	A	T	A	T	A	A	A	T	A	A	A	A	A
6C	cps_>8	T	T	A	A	T	T	A	T	A	T	A	A	A	A	A	A	A	A	A
38	cps_>82	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Due to the number of unexpected amplifications of capsule loci, it was theorised that performing a heat-melt curve when using real-time PCR could distinguish a 'true' amplification based from the positive control.

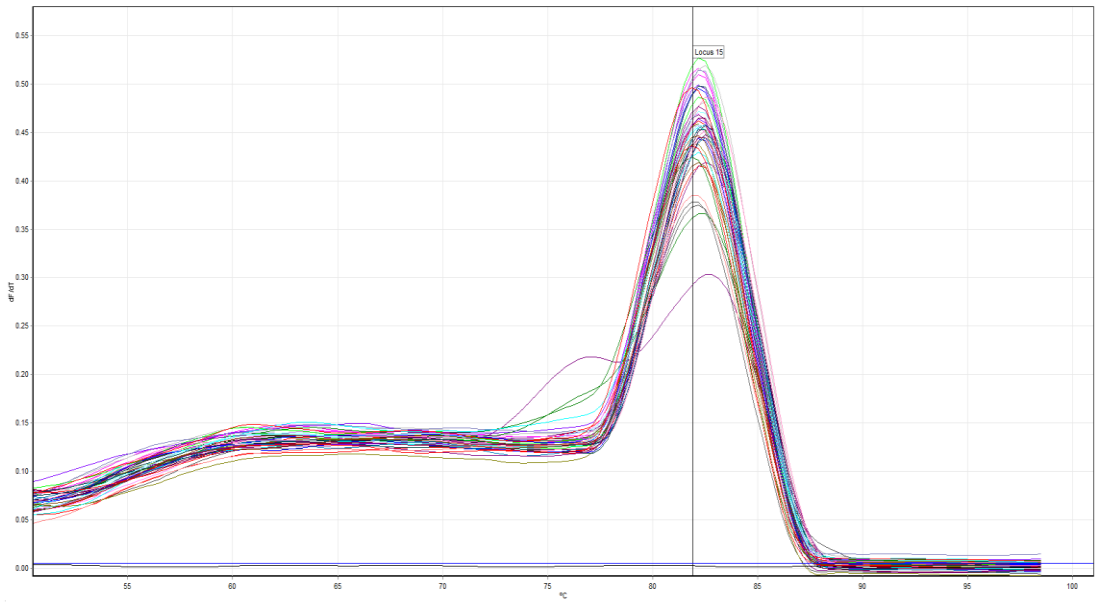
6.3.3 Double-blind validation and real-time PCR

The double-blind validation was performed to determine the capsule types of Queensland pneumococcal isolates. The 48 isolates selected contained at least one isolate from each of the 35 *S. pneumoniae* serotypes detected in Queensland children, and had not previously been capsule typed during the optimisation stages.

Real-time PCR could be used to detect the presence and absence of capsule genes. Each PCR run required a positive control to compare the samples to. For some primers that amplified unexpectedly, we determined that a melt-curve was required when performing real-time PCR to allow the distinction of 'true' present genes and 'false' present genes, depending on the SNPs within the genes. Analysis of the rtPCR results required examination of the Melt A. Green heat curve (Figure 40) which would give clear results of non-amplified genes. A positive control for Locus 15 *rmlB* CPS gene is shown (black line). Amplification of *rmlB* in other pneumococcal isolates was considered 'present' if the bell-curve matched the positive control. For *rmlB*, it is known to be present in a number of pneumococcal capsular polysaccharides, therefore we can see that a number of pneumococcal isolates appear to have a 'present' *rmlB*.

Figure 40: Melt A. Green heat curve for *rmlB* capsule gene (Locus 15). The temperature (°C) is shown on the x-axis.

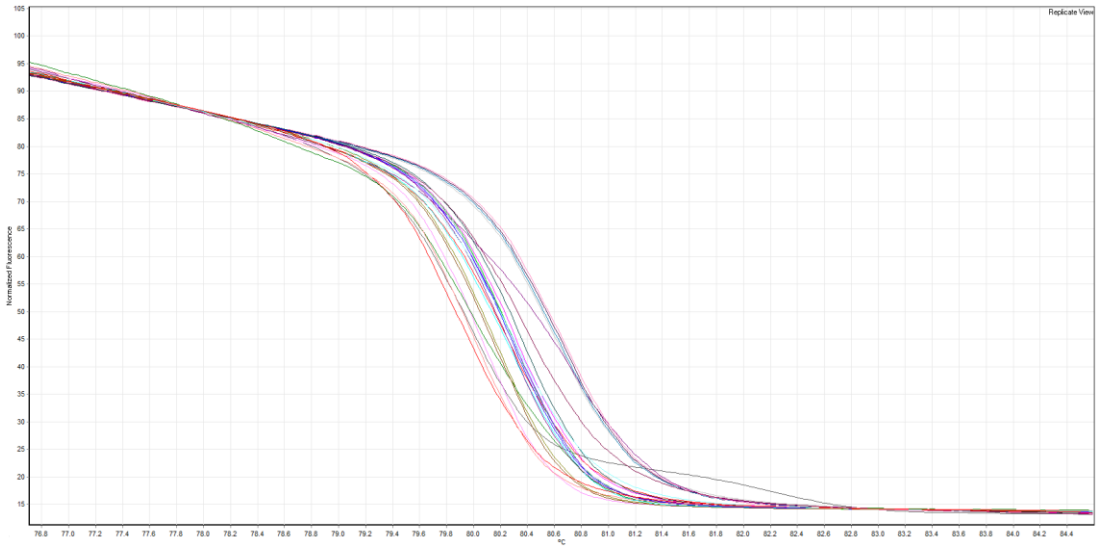
A positive control (Locus 15 – *rmlB*) is shown as the black, vertical line. Each coloured peak is a different isolate with a 'positive' amplification of the gene *rmlB* – if the bell-curve matches the positive control, it is considered a positive amplification and therefore a 'present' gene. An 'absent' gene is observed as a flat line across the graph, or a significantly skewed bell-curve that does not match the positive control.



Following this, analysis of the Normalised High Resolution Melt (HRM) curve would identify whether all positive peaks (in comparison to the positive control present in each run) appeared to follow a consistent curve (Figure 41). Any samples that did not follow the similar curve of the positive control were eliminated (considered 'absent' gene).

Figure 41: Normalised High Resolution Melt (HRM) heat curve for *rmlB* capsule gene (Locus 15).

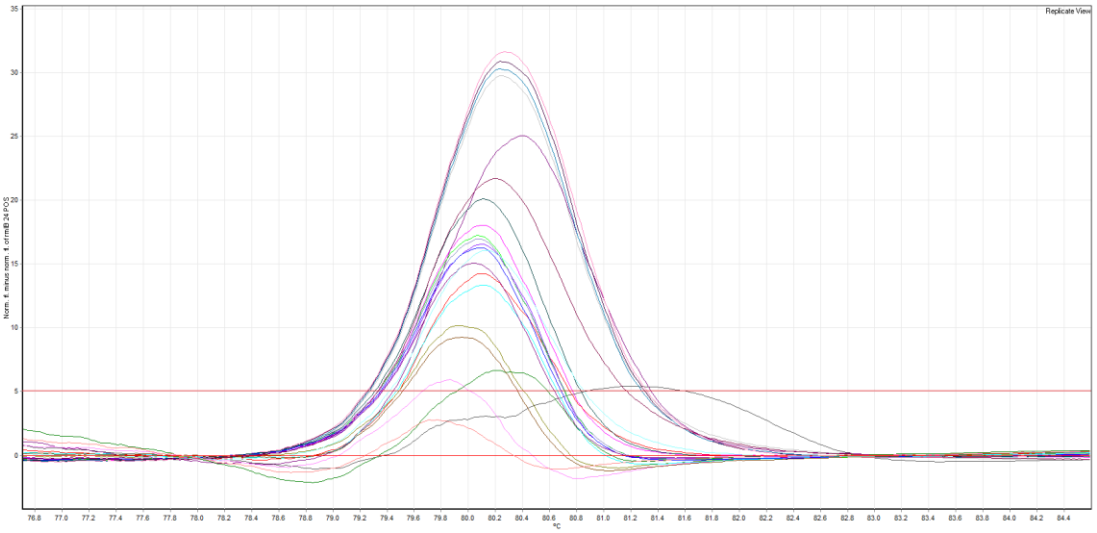
All samples must follow this similar curve pattern to be deemed a 'present' gene (match the positive control – undistinguishable in this graph). 'Absent' genes will follow a more linear, straight line (not shown) or a significantly skewed curve (not shown).



The Normalised HRM is used to create the Difference High Resolution Melt (HRM) - this is then used to determine whether the tested samples closely match the positive control or not as indicated by the presence within ±5 units from the positive control line (Figure 42). The *rmlB* gene is shown as an example and results for the other genes are provided in *Appendix A7: Real-time PCR analysis of 19 primer targets in the* S. pneumoniae *capsule cassette.* The positive control for *rmlB* is set as the base-line '0', and we can set the threshold line (red) at ±5 so that we can easily determine whether the other samples fall outside this threshold. A couple of *S. pneumoniae* isolates have a Difference HRM curve that lies within the threshold and is considered to be definitely 'present', however many other isolates appear to have melt curves that lie outside this threshold.

Figure 42: Difference High Resolution Melt (HRM) heat curve for *rmlB* capsule gene (Locus 15).

A positive control was used for comparison (red line at 0 on y-axis), and peaks that fall outside of the +5 red line were considered false positives.



Unfortunately, as identified before that several genes appeared to amplify unexpectedly because of the primer design, distinguishing the pneumococcal serotypes in this double-blind validation method and using HRM rtPCR was very difficult. Only 13 out of 48 isolates (27%) were correctly identified. This 'correct' identification was based on the known serotype obtained from Quellung serotyping, and the *in silico* expected capsule profile. As a result, we went back to the MinimumSNPs program to identify more genes that could help discriminate the *S. pneumoniae* serotypes.

6.3.4 Re-analysis of Minimum SNPs and double-blind validation using conventional PCR

Re-analysis of the *S. pneumoniae* capsule cassettes using the Minimum SNPs program was performed to identify additional targets to further discriminate the pneumococcal serotypes. A total of 23 CPS genes that was previously selected by MinimumSNPs (D = 0.99) was verified to distinguish between the 93 *S. pneumoniae* capsule cassettes, however 12 capsule genes had not been used previously in this study (*gct, wcwK, wcrH, rbsF, wcwA, wcxM, wcjD, wcrT, wcxD, wcyE, whaK* and *wciY*). It is unknown exactly why approximately half the genes selected by Minimum SNPs are different when analysing the same 182-gene mother template, however it is thought

that these CPS genes are alternatives to provide discrimination, for example we previously described that Binary marker 2 or 5 could be used for discrimination (Figure 34).

Primers were designed for all 23 capsule genes. Some primers were designed to be degenerative primers so that SNP differences within the primer binding site would not affect the ability to amplify the gene (Table 27). Designing primers for *ugd* was difficult as there were no conserved regions within this gene across the *S. pneumoniae* serotypes that contained the gene. This does mean that *ugd* is discriminatory; therefore it was recommended that the Capsule Typing would not initially contain *ugd* and instead would recommend it as a serotype specific gene if required after capsule typing.

Table 27: Re-analysis of the capsule polysaccharide (CPS) genes selected by MinimumSNPs for characterisation of *S. pneumoniae* serotypes.

Forward and reverse primers (e.g. rmlB-F, forward primer; rmlB-R, reverse primer) have length listed (bp). Total expected amplicon length (bp) is shown followed by annealing temperature (Tm) and guanine/cytosine percentage content (GC %). The primers were blasted in NCBI database (yes, Y) and primers that are degenerative are indicated (D). Asterisks (*) against the primer indicates it has been used for the Queensland study.

Gene	Primers	bp	Amp. Length	Tm	GC%	Blast	Deg.
rmlB-F*	CAATGACCATTCTTCAGGAG	20	192	60.2	45	Y	
rmlB-R*	ATCAATCGCATAGCGAAGG	19		63.1	47.4	Y	
glf-F1*	GAGTTAGAATACCGTAGCCTTCG	23	121	62.2	47.8	Y	
glf-F2*	GAACTAGAGTACCGTAGTCTTCG	23		58.3	47.8	Y	
glf-R1*	GYTTGTGTTCGATRATACGAG	21		59.7	42.9	Y	D
glf-R2*	GTTTGTGCTCAATAATGCGAG	21		62.6	42.9	Y	
wchA-F1*	GCTATTTTTGCTCAGACGCG	20	148	65.1	50	Y	
wchA-F2*	GCTATTTTTGCTCAAACTCG	20		59.6	40	Y	
wchA-R1*	TCATCGTCCACCTTAAACAT	20		60	40	Y	
wchA-R2*	TCATTTTCTATCTTAAACAT	20		48.7	20	Y	
wchF-F1*	CTAGTTGTAGGACGTTTTGT	20	325	54.1	40	Y	
wchF-F2*	CTAGTGGTGGGACGATTTGT	20		61.9	50	Y	
wchF-R*	TCHCGGTTAAAYCCAACATC	20		62.4	44.2	Y	D

Gene	Primers	bp	Amp. Length	Tm	GC%	Blast	Deg.
wcjE-F1*	GTCTGTGTATTAGTTGCTTTTG	22	152	56.3	36.4	Y	
wcjE-F2*	GTCTGTATATTAGTTGCGTTTG	22		56.2	36.4	Y	
wcjE-R*	TTTCAACATTAGAATCCGAATCATG	25		64.9	32	Y	
gct-F1	GTACATTTGATTTKCTTCATTATGG	25	265	61	30	Y	D
gct-F2	GAACCTTTGATTTATTGCATTATGG	25		63.4	21	Y	
gct-R1	CCAATCATCCCCCATAACAA	20		64.6	45	Y	D
gct-R2	CCAATCATCACCCATTTCAA	20		63.6	40	Y	
wcwK-F1	ATTGATTTTGTMGTTACATGG	21	178	56.6	31	Y	D
wcwK-F2	ATTGATTTTGTGGTGACATGG	21		62.5	38.1	Y	
wcwK-R1	TTATTAACCCAGGGCGCATGT	21		67.3	47.6	Y	
wcwK-R2	TTATTTACCCAAGGGGCATAT	21		61.2	38.1	Y	
wcrH-F1	TTTGGATTGGTDTGGGATGG	20	173	66.6	46.7	Y	D
wcrH-F2	TTTGGTTTGGTDTGGGAYGG	20		67.6	49.2	Y	D
wcrH-R1	AATACCACAATTATTCTCCT	20		52.4	30	Y	
wcrH-R2	TATACCACACTGCTTATCTA	20		50.4	35	Y	
wcrH-R3	AATTCCAACATTATTTTCTTC	21		54.6	23.8	Y	
rsbF-F*	GGTTGGAAAGGMACGATACAAG	22	220	65.2	47.7	Y	D
rsbF-R*	GGWCCATGATCTGCWACAAA	20		62.2	45	Y	D
wchX-F*	AGATACATATTACGAATTGCCGGATT	26	150	65.3	34.6	Y	
wchX-R*	ACTAGCGGTTCCATTAAAGATTGG	24		65.2	41.7	Y	
wciN-F1*	TTCTTATAAAGAATTATCTCCAA	23	130-160	53.9	21.7	Y	
wciN-F2*	CTCATTTGGTGTACTTCCTCCAG	23		64	47.8	Y	
wciN-F3*	GCGAACTGAAGAACTAATTGAAG	23		61.2	39.1	Y	
wciN-R1*	TTATATKCTTTCYTGCTTCT	20		53.5	30	Y	D
wciN-R2*	CTATACGTTCTCTGGCATTT	20		56.5	40	Y	
wciN-R3*	TCATATATTCCGCAACCGCC	20		66.9	50	Y	
wcwA-F1	CGGGAGAAGAGTGGCGAGTT	20	~180	68.2	60	Y	
wcwA-F2	CTGGTGAAGAGTGGCGAGTT	20		64.8	55	Y	
wcwA-R1	GCCATAGCCTCTAGTACCACTG	22		62	54.5	Y	
wcwA-R2	GCCATCGCCTCTAGTACCACTG	22		67.2	59.1	Y	
wciG-F1*	CAGTTTTGGTTTTTYGGTKC	20	235	61.7	40	Y	D

Gene	Primers	bp	Amp. Length	Tm	GC%	Blast	Deg.
wciG-R1*	AACTGAGATATAAAGCCCCCT	21		60	42.9	Y	
wciG-R2*	AATTGCGCTATAMAACCACCT	21		62.1	40.5	Y	D
wcxM-F*	GGTTGTACGTGGAATCGGAT	20	150	63.9	50	Y	
wcxM-R*	GATTCCTTCTTTTGACAGTGC	21		60.8	42.9	Y	
wcjD-F*	AGGGAGGGATAGATGCAGGTT	21	221	60	52.4	Y	
wcjD-R*	ATCTCCGTACCTGAGAAACCA	21		58	47.6	Y	
wcrT-F1	GCAAAATTAGCTTTAGAGTA	20	~155	51.5	30	Y	
wcrT-F2	GCAGAATTGGCTAAAAAGTA	20		56.8	35	Y	
wcrT-R1	CCAGCTCTCATTAGCCCAAC	20		64.1	55	Y	
wcrT-R2	CCAGTTTTCATTTGCCCAAC	20		64.5	45	Y	
wcjG-F*	GGGATGTGGCAAGTTTCTGGTA	22	150	67.3	50	Y	
wcjG-R*	AATGAGCCCCTTCTCCAGAC	20		64.8	55	Y	
wciK-F*	CGGGAATCACACTCCCTTACTAGA	24	150	66.4	50	Y	
wciK-R*	ATCCTTCATAGGTAAAACGAAACCA	25		64.6	36	Y	
wcxI-F*	AATGATGTGCCTGATACTTTTATTATGC	28	150	64.9	32.1	Y	
wcxI-R*	TTTGCTTAGCTGACACCATTCC	22		65.4	45.5	Y	
wcyE-F	TTGAGGTGCGATTTTTATATTGTGAA	26	150	66.1	30.8	Y	
wcyE-R	TGTTTCCATCTTGCTTCTCCAGTA	24		66	41.7	Y	
whaK-F	AGCAGTATCTTTTGGATGGCAAGT	24	150	66	41.7	Y	
whaK-R	AAGGTCCCATCGCTGACAAT	20		66	50	Y	
wzg-F*	ATTAGTTCGGTGTCGCGATCA	21	150	66.9	47.6	Y	
wzg-R*	TGCCCGCATGGGTTAATTT	19		67.4	47.4	Y	
wciY-F*	TATTTGATAAYTTTGGTGGTAGAGG	25	273	61.3	34	Y	D
wciY-R*	CCATCWGAACCATGCCAWGT	20		65.6	50	Y	D

The selection of targets was minimised to 16 capsule genes for typing Queensland *S. pneumoniae* isolates, including *wzg, rmlB, glf, wchA, wciK, wciN, wcjE, wcjG, wchX, wchF, wciG, rbsF, wcxM, wcjD, wcxI* and *wciY*. This Queensland set has a discriminatory value of 0.9782, although adding *ugd* as an extra gene would increase the discrimination to 0.9886. The other genes not included in the Queensland set would not have added discriminatory value. Additionally, each capsule gene was given a different annealing temperature for optimal amplification, based from the annealing

temperature (Tm) calculated by GeneWorks, Australia when ordering the primers (Table 28).

Table 28: Annealing temperatures (Tm) for the MinimumSNPs targeted loci.

Locus	Tm (°C)	Locus	Tm (°C)
wzg	60	*wchX*	65
rmlB	60	*wciN*	60
glf	61	*wciG*	55
wchA	55	*wcxM*	60
wchF	55	*wcjD*	65
wcjE	61	*wcjG*	60
wcwK	60	*wciK*	60
rbsF	55	*wcxI*	60

A double-blinded capsule typing experiment was performed again, this time using conventional PCR and gel electrophoresis targeting the new 16 capsule genes after the previous re-analysis. Isolates had been re-labelled by someone else other than the person carrying out the experiments. The two *S. pneumoniae* ATCC controls were contained within this double-blinded list of pneumococcal isolates. The double-blind validation improved from the initial rtPCV validation: 21 out of 48 isolates (44%) were correctly identified based on the 16 targeted capsule genes (Table 29). Half of these (11 isolates) were still contained within a complex of other potential serotypes that could have the same profile as well. The amplified genes were visualised using gel electrophoresis (Figure 43 to Figure 56).

A sequential PCR hierarchy table was designed for ease of assigning a serotype based on the amplification of the 16 capsule genes (Figure 57). When simply typing *wzg, rmlB, glf, wchA* and *wchF*, the Queensland *S. pneumoniae* serotypes were divided into eight clusters, of which serotype 1 and serotypes 6B and 6C could already be uniquely identified. After this, amplification of only a few genes could be performed to further differentiate the complex of serotypes (Figure 58 to Figure 61).

Closer examination of whether each individual gene actually amplified in expected serotypes indicated that the majority of CPS genes were good at correctly typing each serotype. Most of the targeted genes would correctly type more than 70% of the 48 isolates, including *rmlB, glf, wcjE, wchA, wchF, rbsF, wcjD, wchX, wciG* and *wciK*. The other genes, *wcxM, wcxI, wciN* and *wcjG* would type less than 60% correctly, but the annealing temperatures were lower for these genes (all 60°C) which could be increased.

Additionally, despite the *in silico* analysis indicating otherwise, several genes appeared to consistently be present or absent in particularly serotypes when they

shouldn't. The gene *rmlB* failed to amplify in serotype 19A when it was expected to be present according to the *in silico* results, but however was present in serotype 4 when it was expected to be absent. Similarly, the *wchF* was absent in all serotype 7F (n=3) and serotype 22F (n=2) when it was expected to be present.

Table 29: Double-blind validation of *S. pneumoniae* using conventional PCR MinSNPs Capsule Typing.

T = present gene (yellow highlight is strong band, pink highlight is faint band and may be incorrect); A = absent gene; V – assigned serotype validated by a second person (X – incorrect assignation, green tick or serotype – correct assignation).

Code	wzg	rmlB	glf	wchA	wchF	wcjD	wcjE	wcjG	rbsF	wchX	wciN	wciG	wcxM	wciK	wcxI	Serotype	V	Expected serotypes
1	T	T	A	T	T	A	A	A	A	A	A	A	T`	A	A	23A/B/C	X	18A
2	T	T`	T	A	A	A	A	A	A	A	A	A	A	A	A	38	X	18B
3	T	T	T	T	A	A	A	A	A	A	T`	A	T`	A	A	6C/B	X	22F
4	T	T	A	T	T	A	A	A	A	T	A	A	T`	A	A	16F/18A	X	23A
5	T	T	A	T	T	A	A	A	A	T	T`	A	T`	A	A	16F/18A	X	23F
6	T	T	A	T	T`	A	A	A	A	A	A	A	A	A	A	19A/F/23A/B/F	23F	23F
7	T	T	T	T	A	A	A	A	A	A	T`	A	T`	A	A	6C/B	X	7F
8	T	A	T`	T	A	T	T	A	A	A	A	A	T`	A	A	9V	√	9V
9	T	T	A	T	A	A	A	A	A	A	T`	A	T`	A	T`	6A/19A/F	X	1
10	T	A	A	T	A	A	T	A	A	A	T`	A	T`	A	T`	9N/11A	11A	11A
11	T	A	T	T`	A	A	A	A	A	A	A	A	A	A	A	38/34	X	19A
12	T	T	T	T	T	A	A	A	A	A	T`	A	T`	A	T`	22A/F/7F/18B/C	18C	18C
13	T	T	T`	T	A	A	A	A	A	A	T`	A	T`	A	T`	6A/19A/F	X	6C
14	T	T	A	T	A	A	A	A	A	A	T`	A	T`	A	T`	6A/19A/F	X	1
15	T	A	T	A	A	A	A	T	A	A	A	A	T`	A	A	10A	X	10F
16	T	A	A	T	A	A	A	T	A	A	T	A	T`	A	A	?	15B	15B
17	T	T	T	T	T	A	A	T	A	A	T	A	T`	A	A	?	22A	22A
18	T	A	T	T	A	A	A	T`	A	A	T`	T	A	A	A	10F/35B/33B/F	33B	33B
19	T	T	A	T	A	A	A	T`	A	A	T	A	A	A	A	6A	√	6A
20	T	T	T`	T	T	A	A	T`	T	A	T`	A	T`	A	T	24F	√	24F
21	T	T	T	T	A	A	A	T`	T`	A	T	A	A	A	T`	6C	X	7F
22	T	A	A	T	A	A	A	T`	T`	A	T	A	T`	A	T`	8/15B/C/14	8	8
23	T	A	A	T	A	A	A	A	T`	A	T	A	A	A	T`	8/15B/C/14	X	12F
24	T	T	A	T	A	A	A	A	T`	A	T	A	T`	A	T`	6A	X	6B
25	T	A	A	T`	A	A	A	A	A	A	T`	A	A	A	A	3/12F	X	19A
26	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	3/12F	X	11A

Code	wzg	rmlB	glf	wchA	wchF	wcjD	wcjE	wcjG	rbsF	wchX	wciN	wciG	wcxM	wciK	wcxI	Serotype	V	Expected serotypes
27	T	T	T	T`	T	A	A	A	A	A	A	A	A	A	A	22A/F/ 7F/18B /C	18C	18C
28	T	A	A	T	A	A	A	A	A	A	T`	A	A	A	A	8/15B/ C/14	X	38
29	T	T	T	T`	A	A	A	A	A	A	A	A	A	A	T`	?	3	3
30	T	A	T	T`	A	A	A	T	T`	A	T`	A	A	A	T`	10F	X	10A
31	T	A	A	T	A	A	A	T	A	A	A	A	A	A	A	3/12F/ 8/15B/ C/14	15B	15B
32	T	A	A	T	A	A	A	T`	A	A	T	A	T`	A	T`	8/15B/ C/14	15C	15C
33	T	A	A	T	A	A	A	T`	T`	A	T	A	T`	A	T`	8/15B/ C/14	15C	15C
34	T	T	T	T`	A	A	A	T`	A	A	A	A	A	A	A	?	19F	19F
35	T	T`	T`	T	T	A	A	T	T`	T	T	T	T`	A	T`	33B	X	23B
36	T	A	T	T	A	A	A	T	T`	A	T	T	A	A	T`	33B/10 F	X	33F
37	T	A	T	T	A	A	T	T`	A	A	T`	T	T`	A	T`	33F	√	33F
38	T	A	T	T	A	A	A	T	A	A	T	T	T`	A	T`	33B	X	35F
39	T	A	T	T	A	A	A	T`	A	A	T	A	T`	A	T`	35B	√	35B
40	T	T	T	T	A	A	A	T	A	A	T	T`	A	A	T`	6C/B	X	7F
41	T	A	A	T	A	A	T	A	A	A	A	A	A	A	A	9N/11 A	9N	9N
42	T	T`	A	T`	A	A	A	T`	A	A	T`	A	A	T	T`	4	√	4
43	T	A	A	T	A	A	A	A	A	A	T`	A	A	A	T`	8/15B/ C/14	14	14
44	T	A	T	T	A	A	A	A	A	A	T`	A	A	A	A	34	√	34
45	T	T	A	T	T	A	A	T`	A	T	T`	A	T	A	T`	16F	√	16F
46	T	T`	T	T	A	A	A	T`	A	A	T`	A	T	A	T`	6C/B	X	22F
47	T	T	A	T	A	A	A	A	A	A	T`	A	A	A	A	19A/F ATCC4 9619	ATCC 4961 9	ATCC 4961 9
48	T	A	A	T	A	A	A	T`	A	A	A	T	T`	A	A	ATCC6 306	√	ATCC 6306
NEG control	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	NTC		
S. agalactiae	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	NON S.PNEUMO		

Figure 43: Double-blind validation of *rmlB* on *S. pneumoniae* isolates.

Marker ladder MVIII is in lane 1 in every row; gene name is provided above the first pneumococcal isolate (#1); 1-48 are the coded numbers assigned to each pneumococcal isolate; -ve is negative water control; *S. aga* is short for *S. agalactiae* isolate tested.

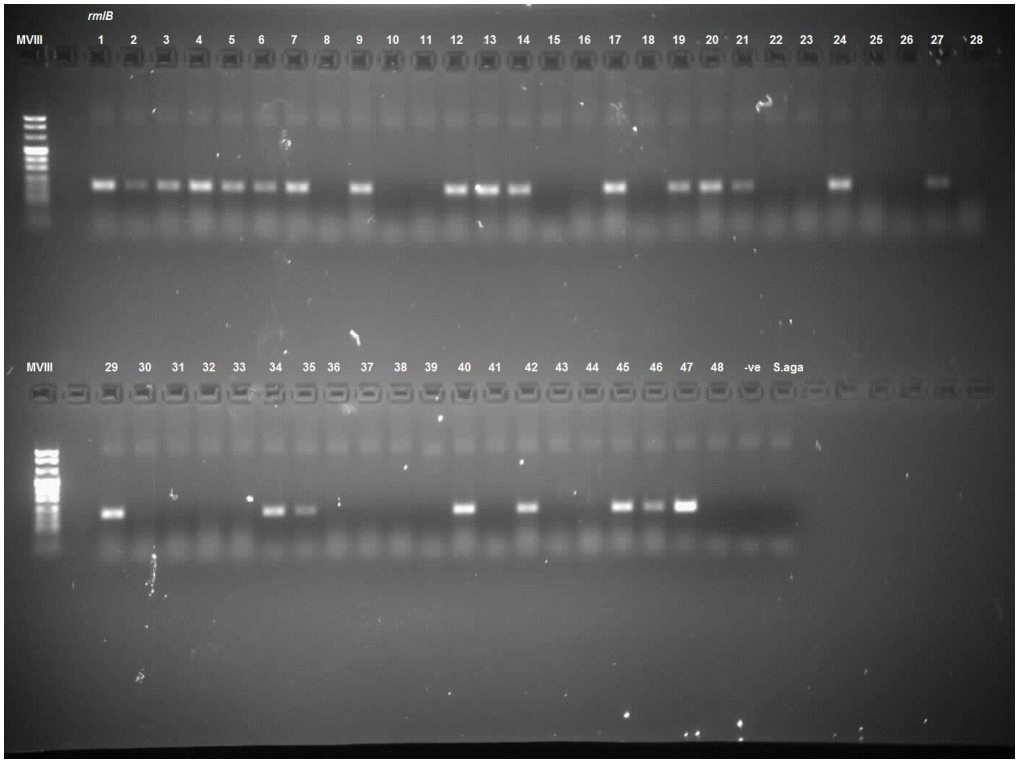


Figure 44: Double-blind validation of *glf* on *S. pneumoniae* isolates.

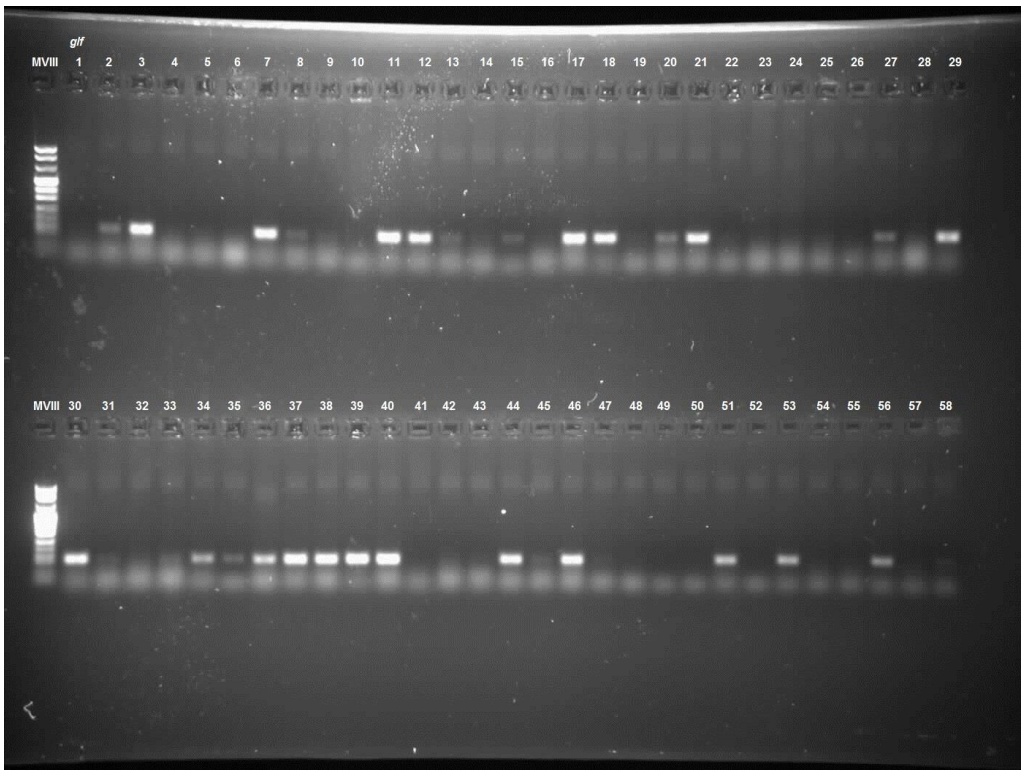


Figure 45: Double-blind validation of *wchA* on *S. pneumoniae* isolates.

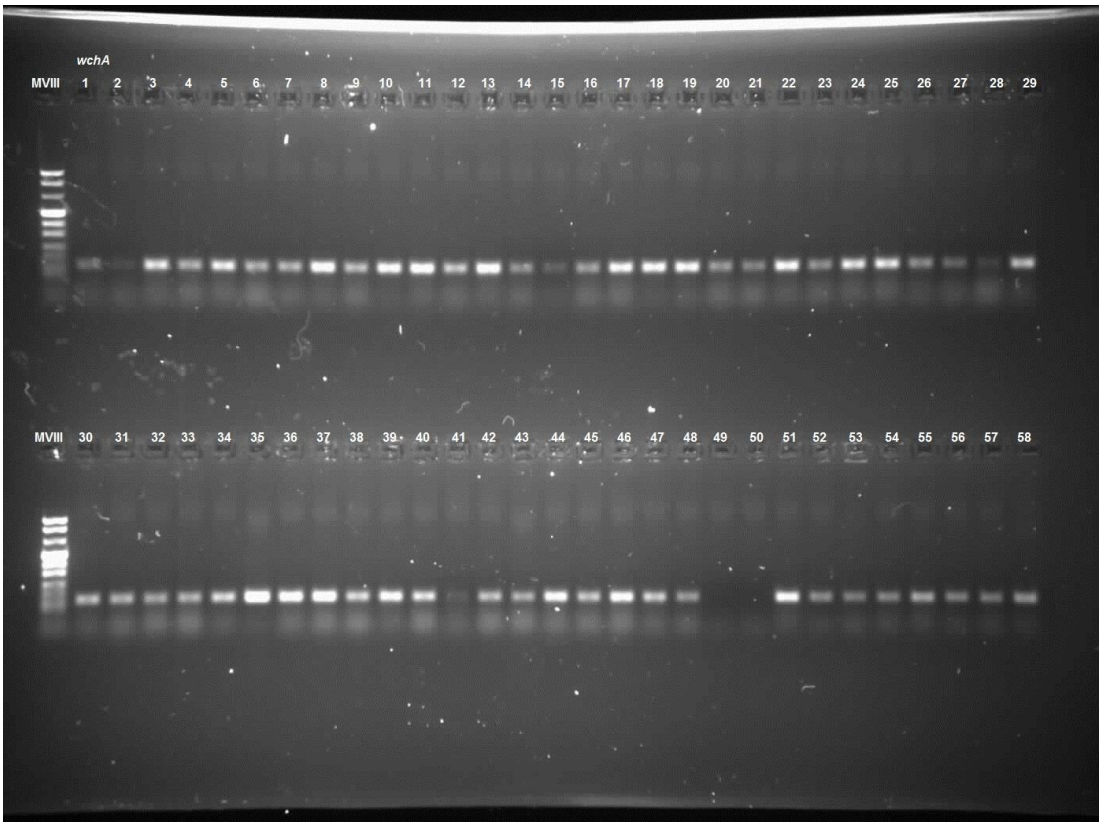


Figure 46: Double-blind validation of *wchF* on *S. pneumoniae* isolates.

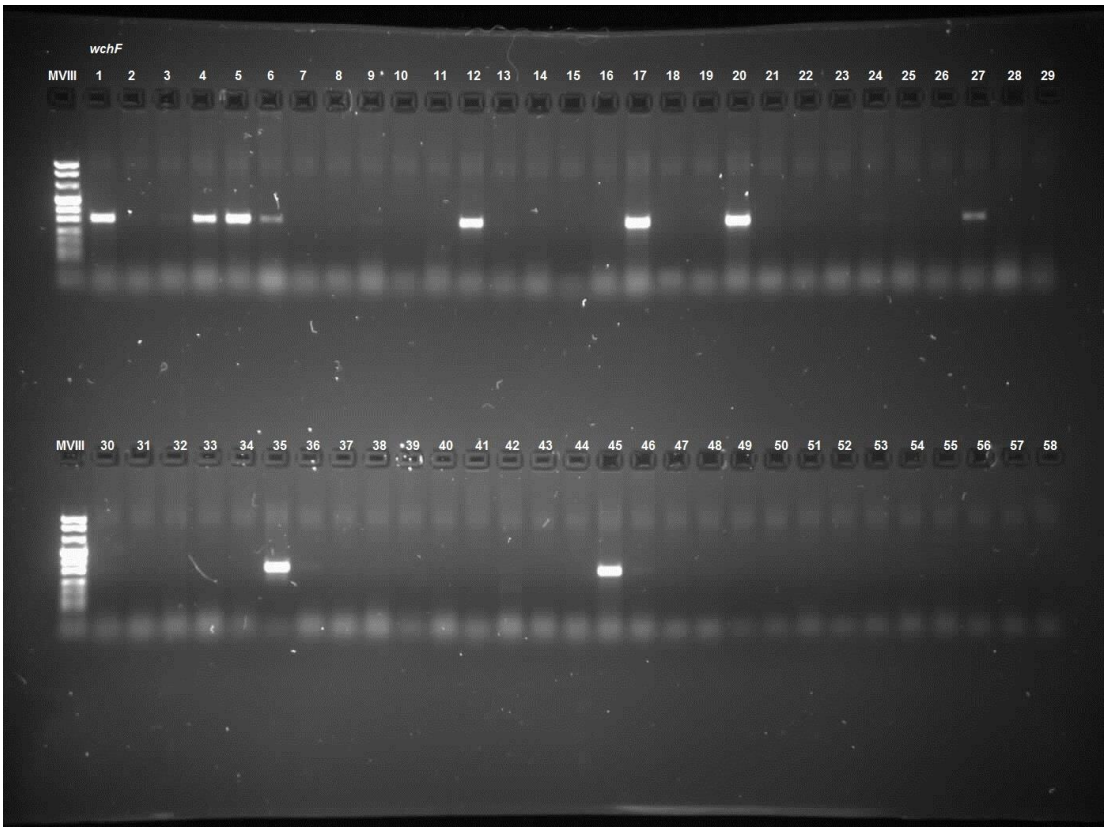


Figure 47: Double-blind validation of *wcjD* on *S. pneumoniae* isolates.

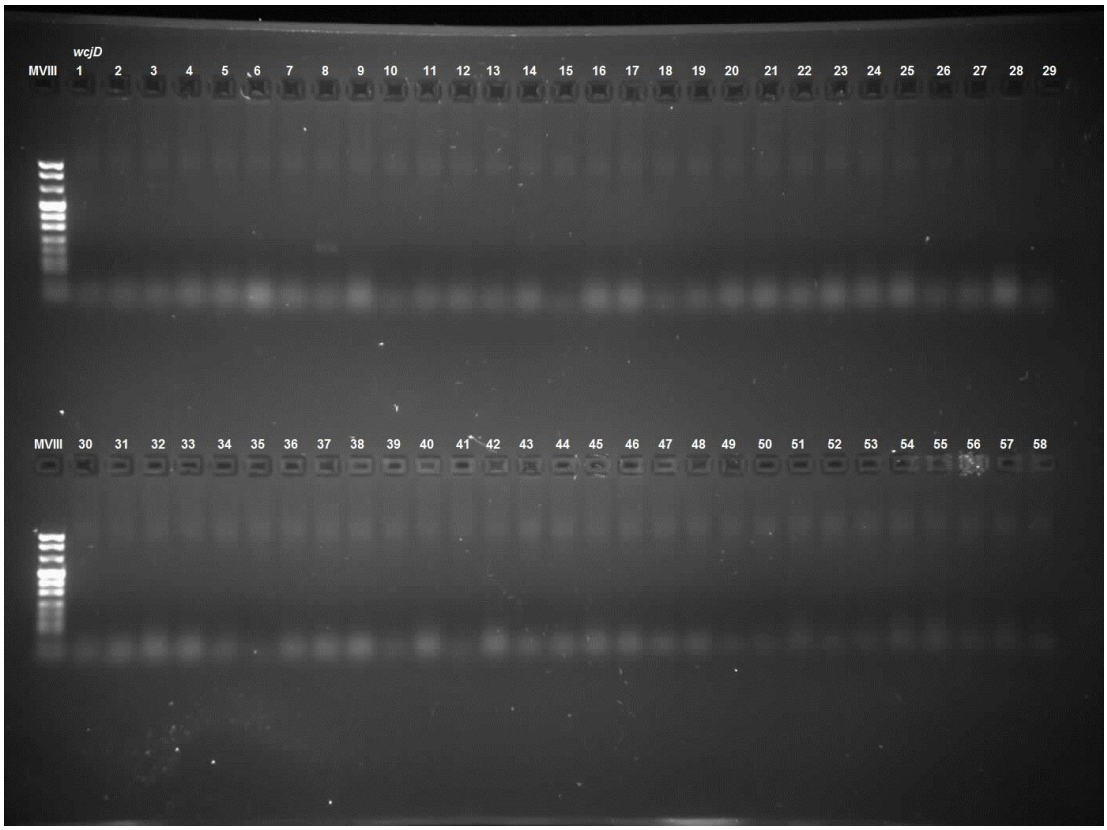


Figure 48: Double-blind validation of *wcjE* on *S. pneumoniae* isolates.

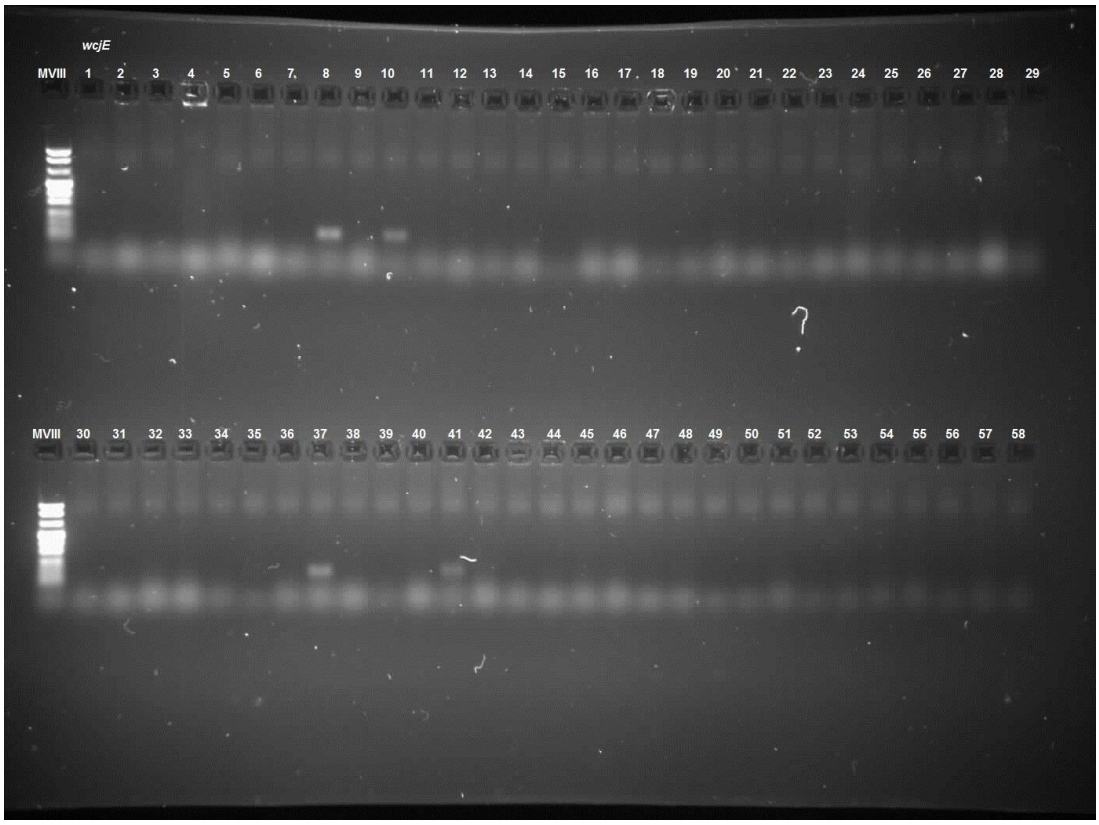


Figure 49: Double-blind validation of *wcjG* on *S. pneumoniae* isolates.

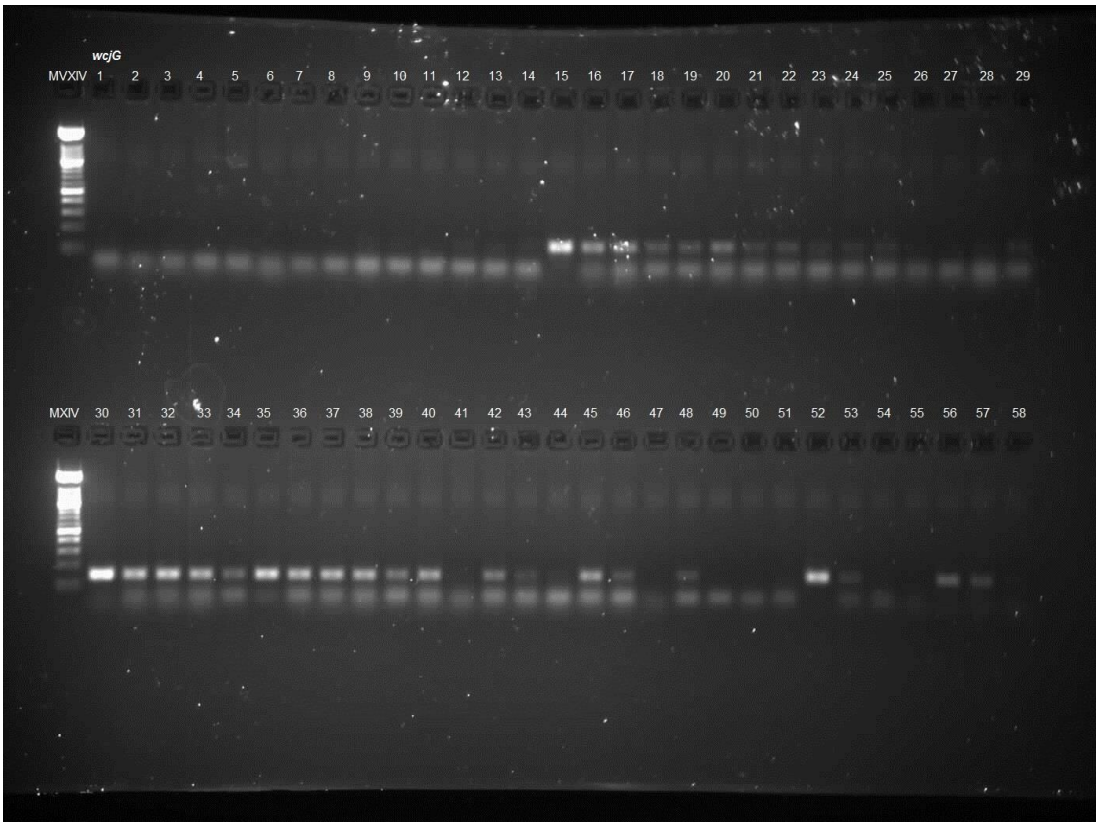


Figure 50: Double-blind validation of *rbsF* on *S. pneumoniae* isolates.

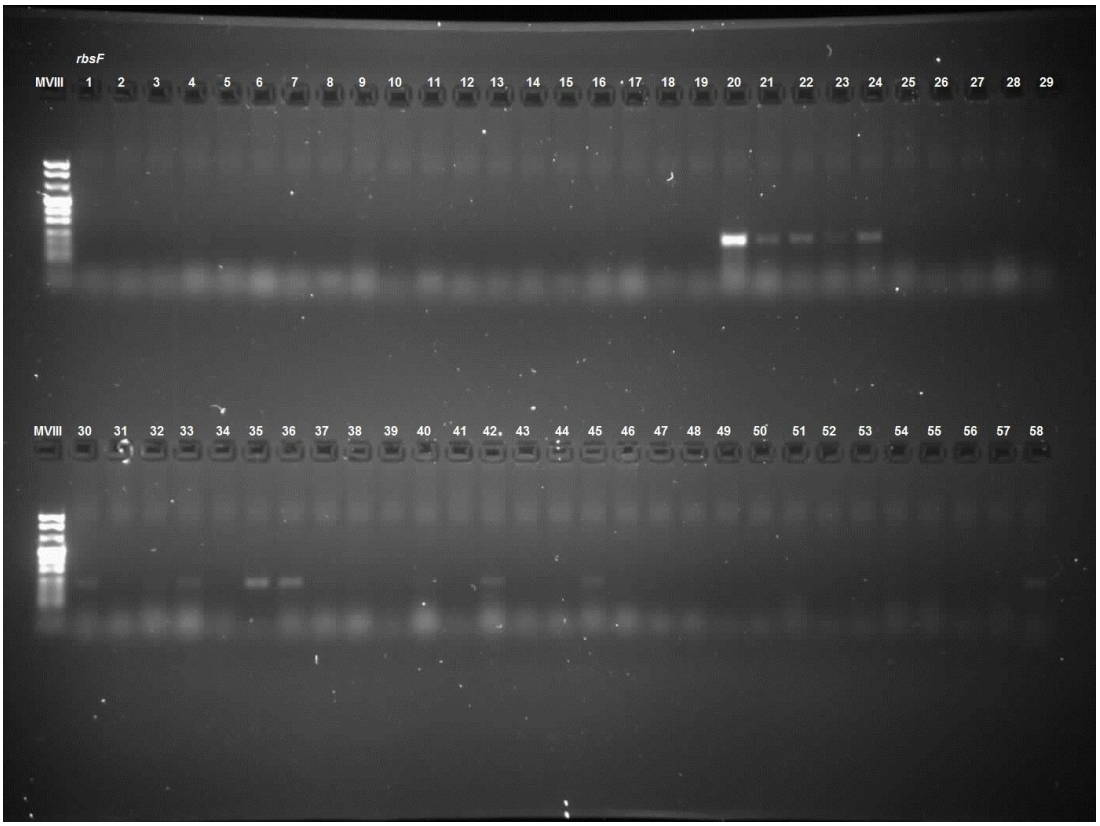


Figure 51: Double-blind validation of *wchX* on *S. pneumoniae* isolates.

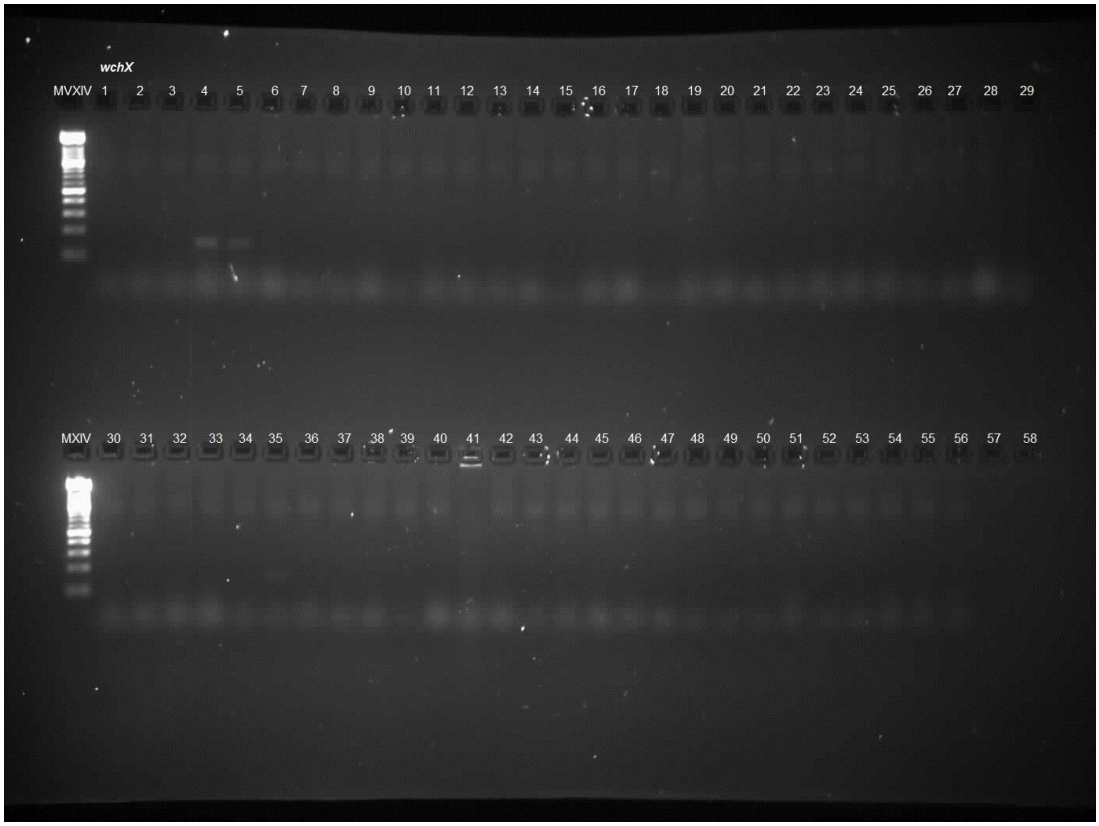


Figure 52: Double-blind validation of *wciN* in *S. pneumoniae* isolates.

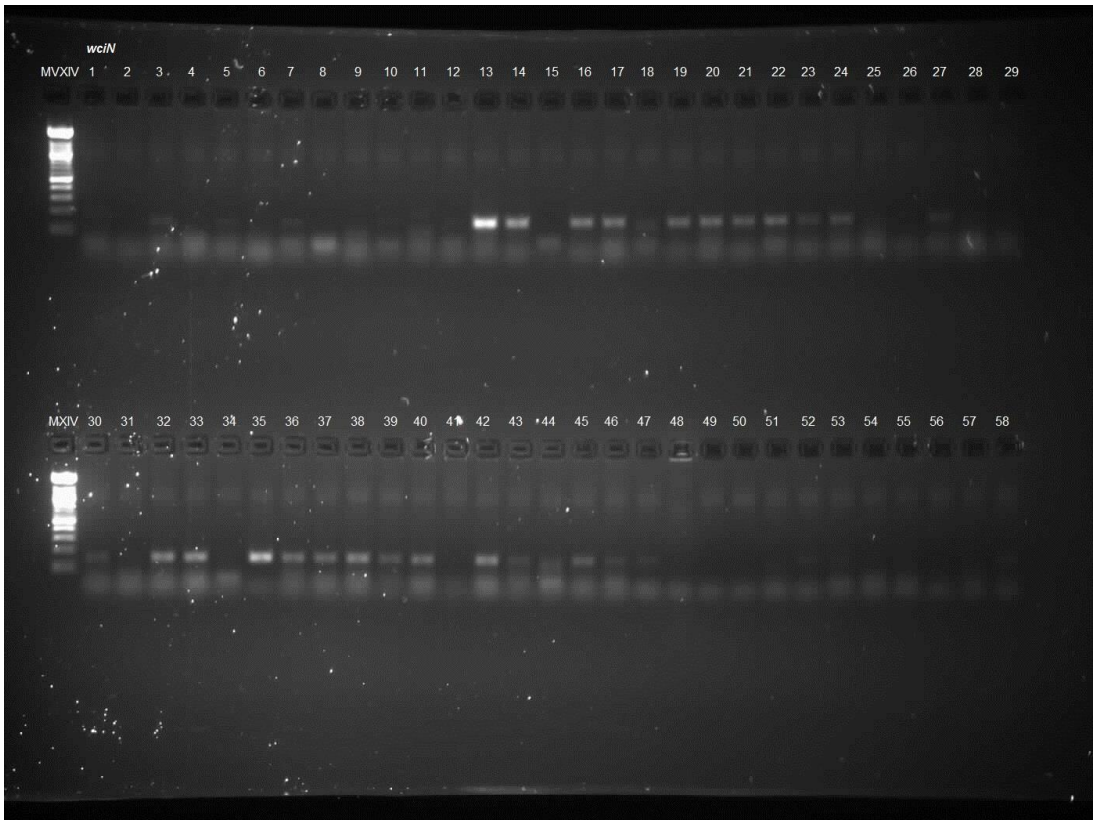


Figure 53: Double-blind validation of *wciG* in *S. pneumoniae* isolates.

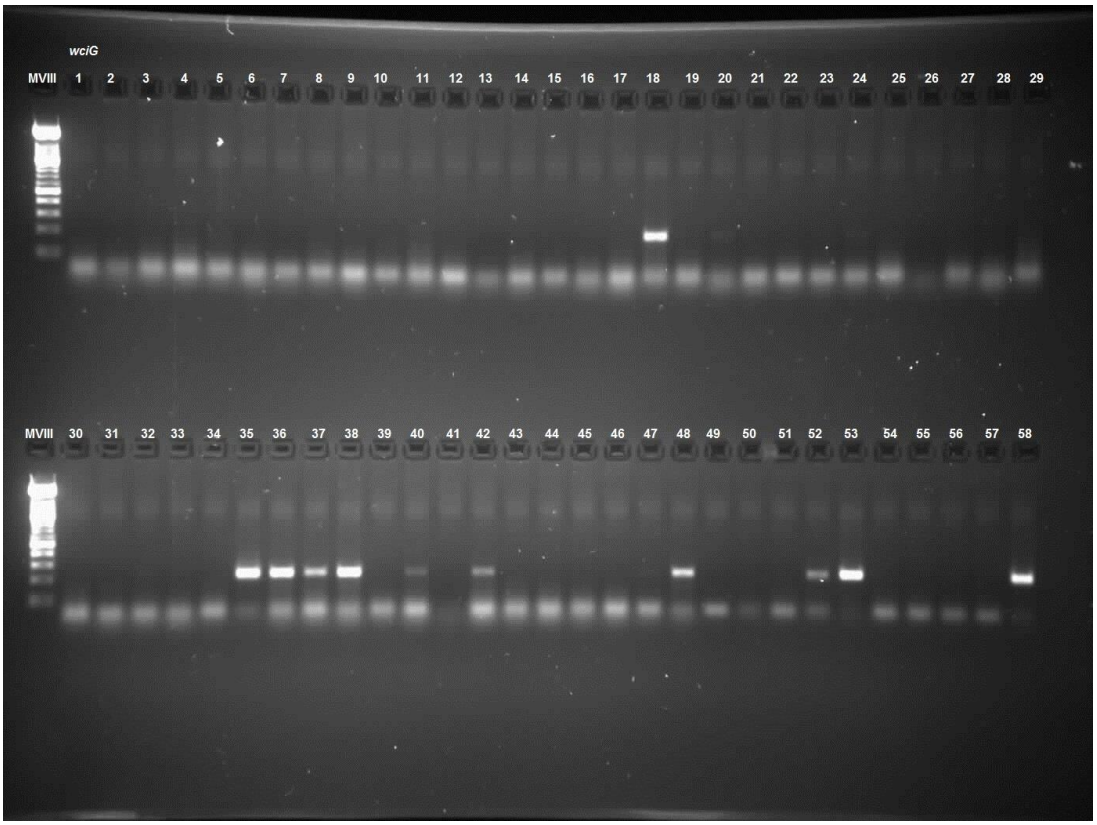


Figure 54: Double-blind validation of *wcxM* in *S. pneumoniae* isolates.

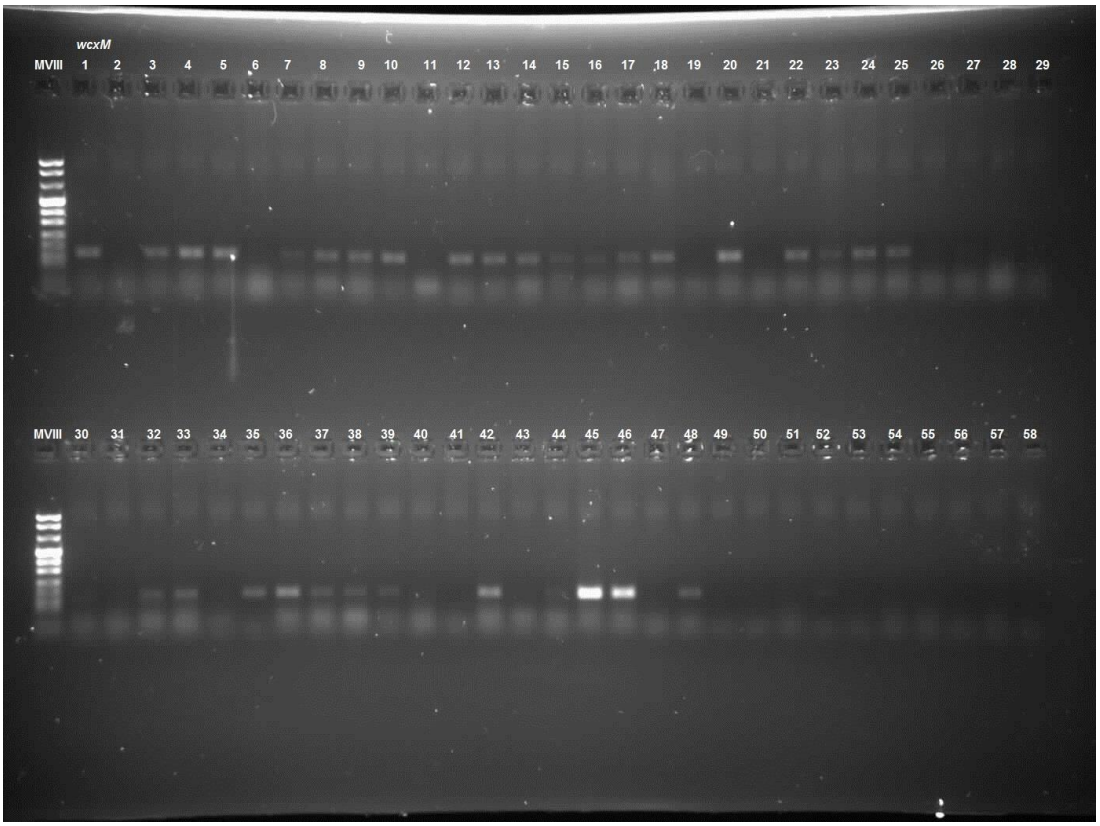


Figure 55: Double-blind validation of *wciK* in *S. pneumoniae* isolates.

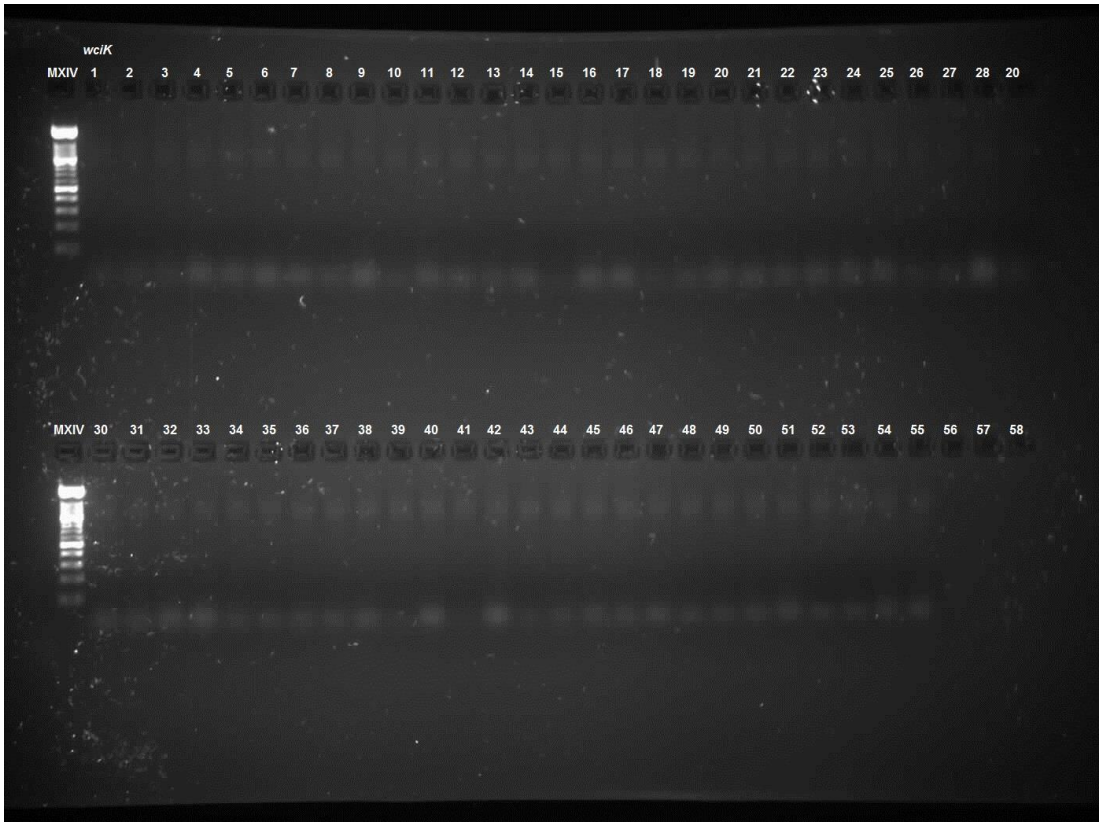


Figure 56: Double-blind validation of *wcxI* in *S. pneumoniae* isolates.

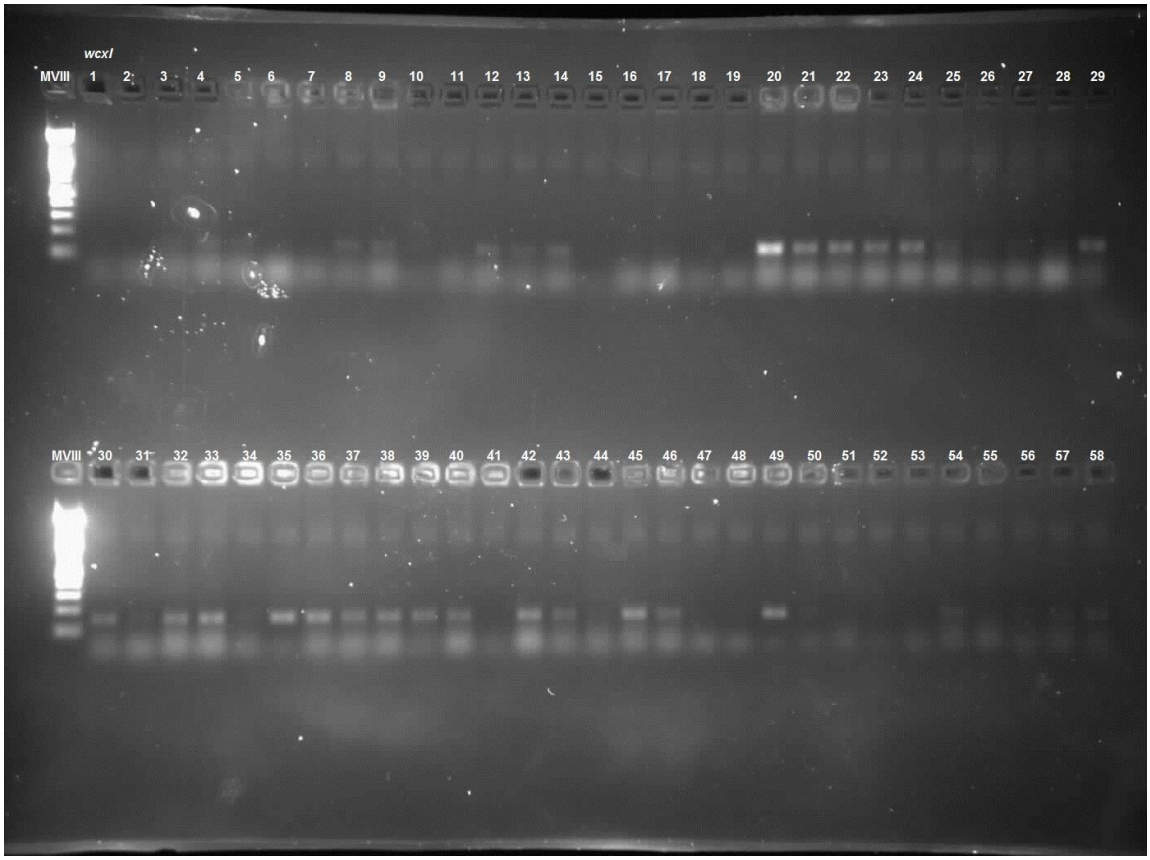


Figure 57: Sequential PCR MinSNP Capsule Typing using *wzg, rmlB, glf, wchA* and *wchF* in *S. pneumoniae.*

Red boxes indicate final serotype that can be given to the *S. pneumoniae* isolate. The bottom 1, 2, 3 and 4 links to consecutive hierarchy figures.

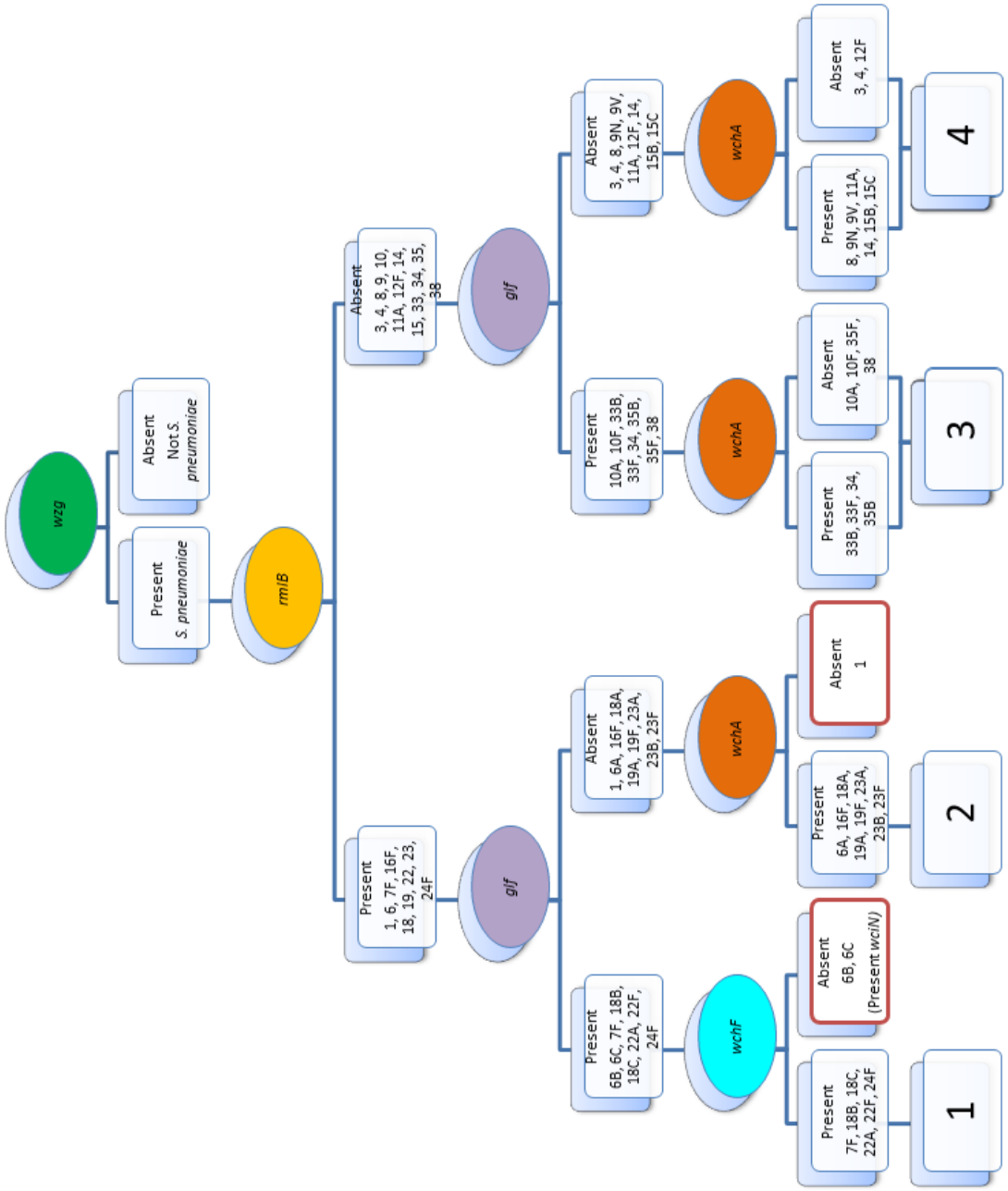


Figure 58: Consecutive hierarchy number 1 for *S. pneumoniae* isolates.

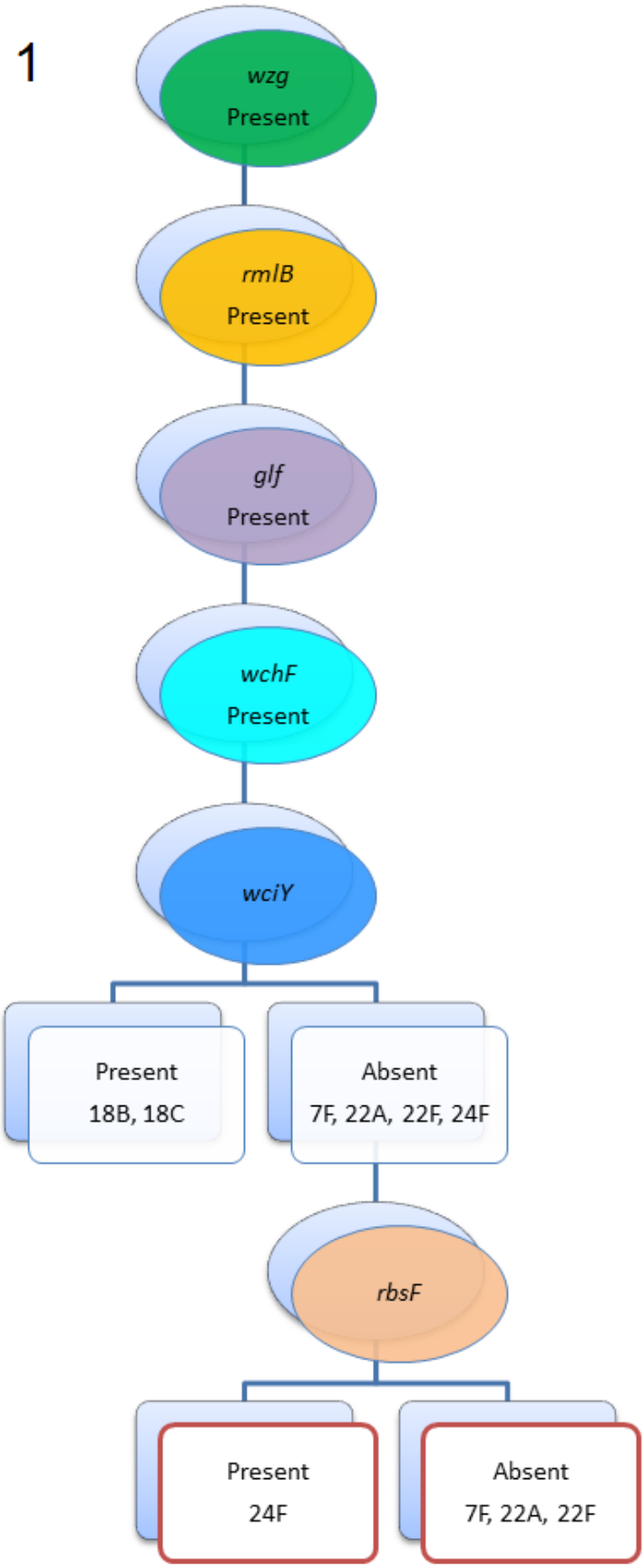


Figure 59: Consecutive hierarchy number 2 for *S. pneumoniae* isolates.

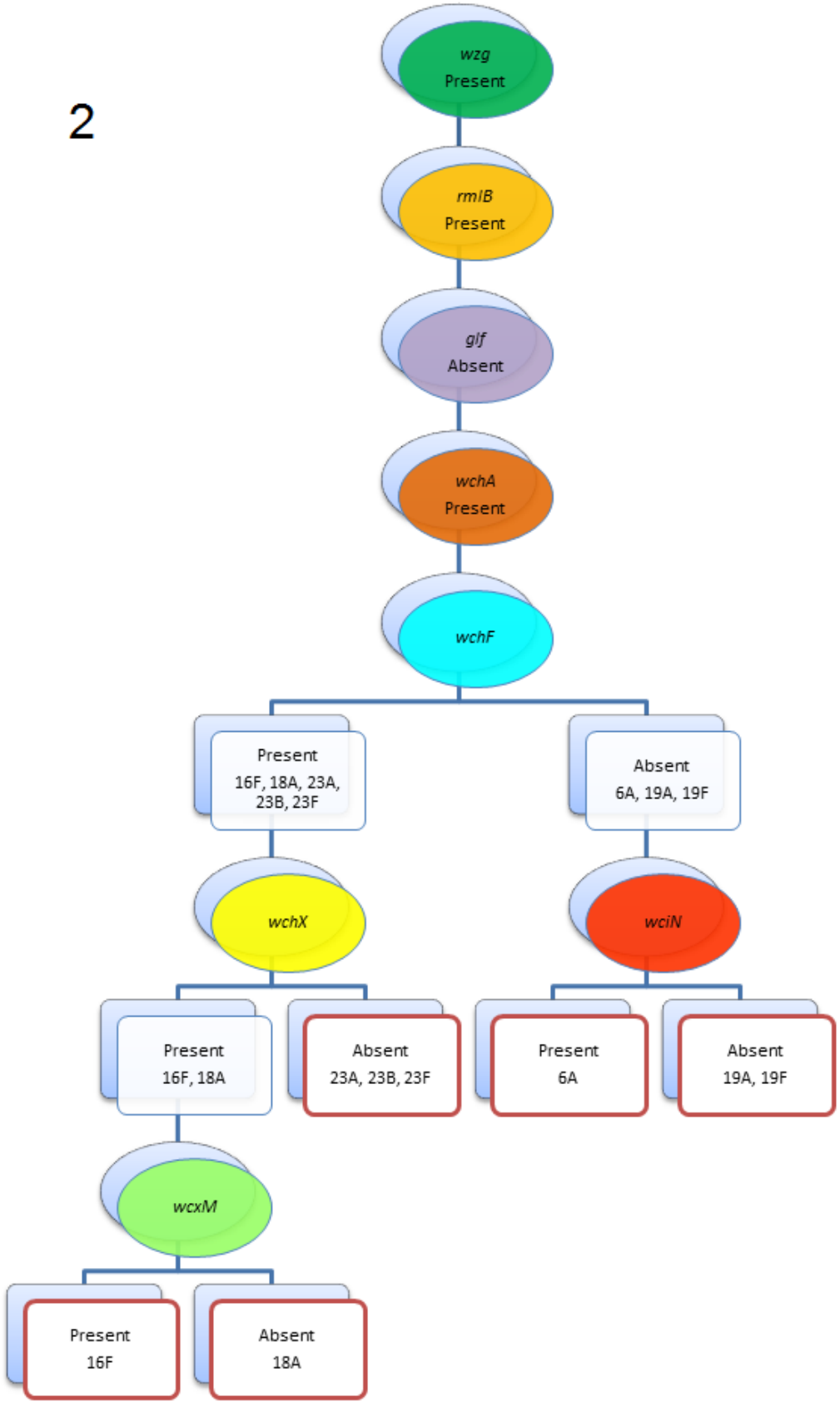


Figure 60: Consecutive hierarchy number 3 for *S. pneumoniae* isolates.

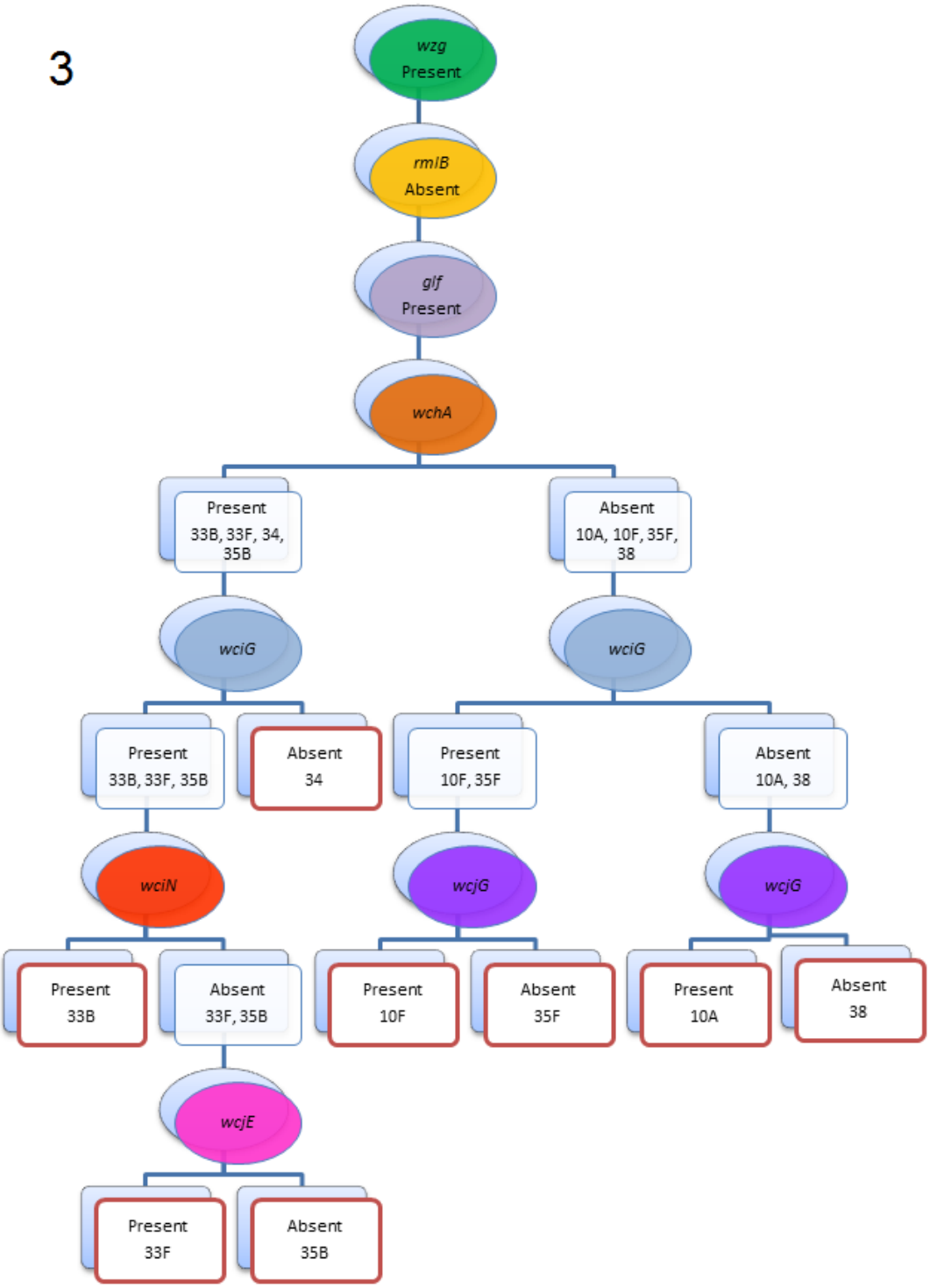
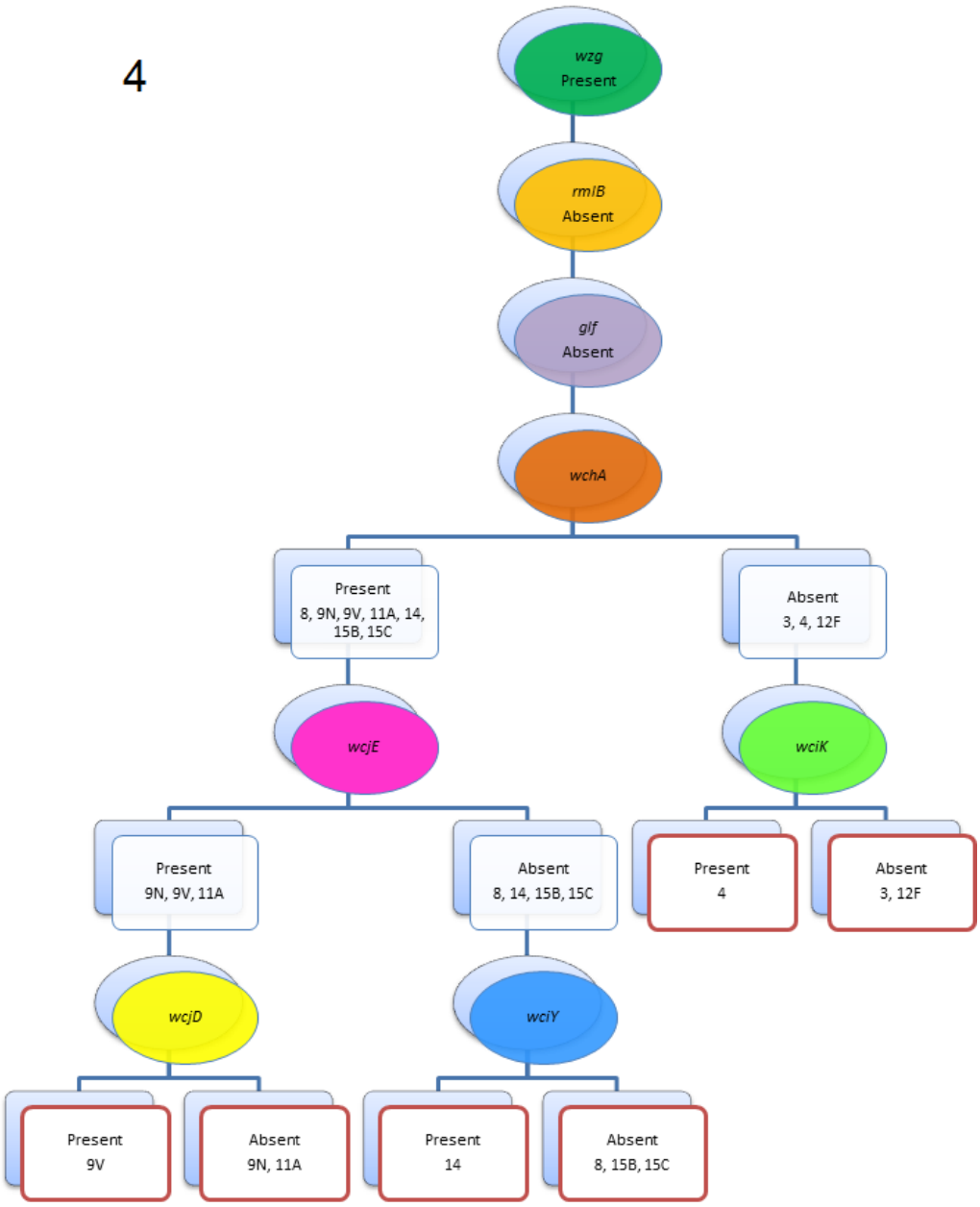


Figure 61: Consecutive hierarchy number 4 for *S. pneumoniae* isolates.



6.3.5 Validation of capsule switching using the novel capsule typing and MLVA4

In addition to capsule typing of *S. pneumoniae*, the novel capsule typing was combined with MLVA4 (previously described in Chapter 4) to determine whether capsule switching events had occurred in Queensland. From the previous study (Chapter 4) and combining traditional serotyping with MLVA4, two potential capsule switches had been identified, including a switch from serotype 1 to serotype 4, and a

switch from serotype 19A to serotype 15C. The capsular profiles based on the novel capsule typing have verified that the capsular profiles for these respective *S. pneumoniae* isolates indicate that even though the isolates have the same genotype, they have different capsule types (Table 30). Using the novel capsule typing, this determined that a potential capsule switch had occurred in Queensland pneumococcal population. Examination of the MLST database, which contains the most comprehensive data of *S. pneumoniae* strains worldwide, can be used to determine whether the capsular switch has more likely occurred in Queensland or not, for example, the result of importation of the strain already "capsule switched" elsewhere. From the database, ST306 (n=251) has been observed mainly in serotype 1 (n=244), however also serotype 9V (n=2), serogroup 11 (n=2), serotype 14 (n=1) and serotype 7F (n=1). There have been no reports of serotype 4 (ST306) elsewhere in the world; only Queensland. This could possibly indicate that the potential capsule switch from serotype 1 to serotype 4 may have occurred in Australia.

Similarly, ST411 (n=17) has been mainly recorded in Australia (MLST database) and associated with serotypes 19A, 15B and 15C – in the UK a serotype 15B (ST411; n=1) has been observed, and serotype 15C observed in China (n=1). The serogroup 15 may have been introduced to Australia from overseas; however the capsule switch between serogroup 15 and serotype 19A may have occurred in Australia, since there is no record of ST411 serotype 19A elsewhere in the world. However the low number of isolates submitted to the MLST international database (n=17) limits our confidence in determining this capsule switch.

Finally, ST53 (n=51), according to MLST database, is mainly associated with serotype 8. Only a serotype 9V (n=1) and a non-typeable (NT; n=1) has been reported with ST53. Therefore the serotype 11A, ST53 observed in this Queensland study may most likely be due to a switch within Queensland.

Table 30: Capsule profiles observed for *S. pneumoniae* isolates that may have undergone capsule switching.

Isolate ID	wzg	rmlB	glf	wchA	wchF	wcjD	wcjE	wcjG	rbsF	wchX	wciN	wciG	wcxM	wciK	wcxI	Serotype	Genotype
36	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	1	ST306
37	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	4	
59	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	19A	ST411
172	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	15B/C	MT59
284	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	15B/C	
47	T	A	A	T	A	A	T	A	A	A	A	A	A	A	A	11A	ST53
233	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	8	

6.3.6 Costs of MinSNP Capsule Typing vs. serotyping

Comparison of the cost of traditional serotyping compared to the developed MinSNP Capsule Typing was performed. To obtain a complete set of antisera through Australian agent for the traditional serotyping of any *S. pneumoniae* serotype costs AU$35, 712.00 (May 2014 est.) to completely type and factor type the isolates (H. Smith, personal communication, 30 October, 2014). Multiple tests can be performed. On the other hand, *S. pneumoniae* isolates can be directly sent to Statens Serum Institute, Denmark, who will charge AU$325.00 per isolate to obtain a complete type and factor type.

In comparison to the serotyping costs, it was estimated that Minimum SNPs capsule typing using real-time PCR would cost approximately AU$27.00 per isolate if using the 19 target primer set. This cost includes the SybrGreen master mix reaction (AU$1390.00 for 1000 reactions) and primers (~AU$15.00 each).

6.4 Discussion

Capsule typing of pneumococci has become important in epidemiology studies worldwide, particularly since the polysaccharide capsule plays a vital role in virulence and hence is the target of several childhood vaccines. For decades, a traditional serotyping method based on antisera reaction has been widely used in epidemiology studies, however a number of limitations of this method have prompted the development of molecular based capsule typing methods. Of particular interest is the amplification of the capsule cassette, or discriminatory genes within the capsule cassette to characterise the more than 90 different pneumococcal serotypes. Currently, no molecular typing method has characterised all 98 serotypes, therefore this study has focused on a novel approach to select a minimum set of highly discriminatory capsule genes that have the ability to characterise all 98 serotypes. Furthermore, by determining the capsule type of each pneumococcal isolate and combining this information with the genotype (either MLST or MLVA), verification of whether capsule switching has occurred in the Queensland population. It was hypothesised that the novel capsule typing can detect *S. pneumoniae* capsule types and sub-types and that capsule switching events can be verified by combination of the novel capsule typing with MLVA4.

A Minimum SNPs software program was used to identify the minimum set of highly discriminatory capsule genes within the pneumococcal capsule cassette. This software program has previously been used by several studies to identify highly

discriminatory SNPs for characterisation of bacteria (Price *et al.,* 2006; Rathnayake *et al.,* 2011; Robertson *et al.,* 2004; Sheludchenko *et al.,* 2010; Stephens *et al.,* 2007). In this study, we were able to utilise it to identify highly discriminatory genes (rather than SNPs) by developing a pseudoDNA sequence that consists of only A (absent gene) and T (present gene). The Minimum SNPs program was then able to identify the highly discriminatory 'SNP' which codes for a particular gene, demonstrating that this bioinformatics program can be utilised not only to detect SNPs but also to detect discriminatory genes within a cassette.

In total, 23 capsule genes were identified that would give the highest discrimination between all 93 available sequences of pneumococcal capsule types. However, more than one primer pair had to be designed for some of these genes because it proved difficult to develop a universal primer pair to target all expected serotypes. Some primers were also degenerative primers to allow amplification of SNP differences between binding sites. A number of the selected genes have been used in other capsule typing studies, including *ugd* (Tarrago *et al.,* 2008), *wchF* (Tomita *et al.,* 2011), *wcjE* (Bentley *et al.,* 2006; Calix & Nahm, 2010; Calix *et al.,* 2012), *wciN* (Bratcher *et al.,* 2011; Carvalho *et al.,* 2009; Jin *et al.,* 2009; Park *et al.,* 2007; Pimenta *et al.,* 2013; Tomita *et al.,* 2011) and *cpsA* or *wzg* for internal control. This demonstrates that these genes have been recognised as suitable capsule typing genes, and that the Minimum SNPs program has similarly identified them to be included in this scheme. The *wzg* failed to amplify in *S. agalactiae* and *S. pyogenes* as expected and can be used to distinguish *S. pneumoniae* from other streptococcal species. However, other pneumococcal related species, such as *S. oralis, S. mitis* and *S. pseudopneumoniae*, could have been tested, particularly since a study has demonstrated that non- *S. pneumoniae* species can confound serotype-specific PCR due to evolutionary conserved genes (Carvalho *et al.,* 2013). This could be a potential limitation of the novel capsular typing method. Further studies would include validating the capsule typing method on other pneumococcal species, particularly *S. mitis.* The *wzg* is contained in a homologous region at the 5' end of the capsule cassette (Bentley *et al.,* 2006), and was consistently amplified in all *S. pneumoniae* isolates in this study.

For capsule typing the 35 known serotypes found in Queensland, only 16 gene targets were required that gave a Simpson's Index of Diversity of 0.9782, of which further genes or SNP-specific primers could then be utilised to further differentiate any complex of serotypes that hadn't been differentiated. Several genes unique for differentiation of *S. pneumoniae* serotypes were observed in this study. For example, *wciN* has been used to distinguish between serotype 6A and 6B (Bratcher *et al.,* 2011),

however this study utilised *glf* for differentiation. The insertion of *wcjD* in serotype 9A and 9V differentiate it from serotype 9L and 9N (Bentley *et al.,* 2006). This study observed that serotype 9V had a present *wcjD* whereas serotype 9N did not. Another gene *wcjE* that is known to be mutated in serotype 9N was still observed to be present in this study (Bentley *et al.,* 2006). Similarly, *wcjE* can be used to differentiate serotype 11A from 11E (Calix *et al.,* 2010), however no serotype 11E was observed in this study to verify this, although *wcjE* was present in serotype 11A. Serotypes 10C and 10F can be differentiated from serotypes 10A and 10B by the presence of *wciG* in the former serotypes (Mavroidi *et al.,* 2007); however this was not observed in this study as both serotypes 10A and 10F appeared to lack *wciG*. Finally, *rbsF* can be used to distinguish serotype 24A from 24F, and this study observed that serotype 24F has a present *rbsF* (Mavroidi *et al.,* 2007). The verification of these genes in practice has reassured the *in silico* analysis when differentiating between *S. pneumoniae* serotypes.

Even though 44% is low typeability, further examination of the individual targeted loci indicate that further optimisation of the PCR method may improve the typeability results. Using conventional PCR and gel electrophoresis improved typeability compared to the rtPCR HRM, however this is thought to be because we re-analysed the mother template using the Minimum SNPs program, and it identified 12 new CPS targets that could be used instead for differentiation. As a result, new primers were designed, and to improve PCR conditions each primer pair was amplified at the optimum annealing temperature i.e. there wasn't one set temperature for all primers, instead some primers were amplified at 60^{o}C while others were amplified at 65^{o}C. Due to time and limited funds, we couldn't repeat the experiment with the newly designed primers on rtPCR HRM. Several genes could have a more stringent PCR reaction by increasing the annealing temperature (Tm) so that faint bands that were amplify would disappear, leaving only the 'true' present gene. The G/C content of *S. pneumoniae* capsule cassette is lower than the rest of the genome (Bentley *et al.,* 2006). This may have made it difficult when designing suitable primers with optimal G/C content.

Additionally, several targeted capsule genes appeared to be unexpectedly present or absent in specific serotypes. Unfortunately for most serotypes there are only one capsule sequence available in the NCBI database, therefore it is possible that the Queensland isolates have variations which may indicate that the primers will anneal, or that the capsule genes are present/absent in contrast to the published sequences. At this point none of these capsule genes have been sequenced. It would be of interest to examine closely the following genes that this study noticed to be present or absent in a specific serotype: *rmlB* was absent in serotype 19A (should be present), and present in

serotype 4 (should be absent), and *wchF* was absent in serotype 7F and serotype 22F (should be present in both serotypes). Since the double-blind validation only tested at most 3 isolates per serotype, further validation would be ideal to determine if other isolates of the same serotype produce the same results.

For ease of determine the serotypes from the capsule typing profile, a sequential hierarchy diagram was developed. From this only five main genes are necessary to be amplified to divide the Queensland serotypes into eight groups, and from which only a couple of genes would then differentiate the remainder. This sequential hierarchy diagram has only been designed for the identification of Queensland *S. pneumoniae* serotypes, however the principle can be applied for typing all 98 known serotypes worldwide.

The ability to use real-time PCR with a high resolution melt curve has decreased time and cost compared to traditional serotyping and conventional PCR. Further optimisation is required for the novel capsule typing as only 27% of isolates were correctly identified, however this was only when using the original 19 CPS genes amplified at an annealing temperature of 67^{o}C; as we explained earlier, this typeability was improved to 44% after re-analysing the CPS genes and modifying the PCR conditions so that the new 16 CPS targets would be amplified at different annealing temperatures. We were unable to repeat the second CPS typing with 16 genes on rtPCR HRM due to time and limited funding. The application of rtPCR has been demonstrated for developing a novel capsular typing method for *S. pneumoniae*, although further validation and modification of the method is required to obtain a useable method for routine diagnostic detection.

Results, using rtPCR, could be obtained in minutes since it is not necessary to wait through full amplification which is important in outbreak situations (Bricker, 2011). It was estimated that this real-time PCR cost on average AU$27.00 to identify each capsule type. In comparison, traditional serotyping with a complete set of antisera would cost AU$35, 712.00, or alternatively isolates sent to Statens Serum Institute, Denmark, would cost AU$325.00 per isolate. It is clear that using molecular methods for capsule typing *S. pneumoniae* is dramatically cheaper than conventional serotyping. Further studies would also need to compare other capsular typing methods, also based on PCR (*Appendix A4: Comparison of various pneumococcal PCR-based serotyping methods*).

The second objective of Aim 4 was to determine whether capsule switching could be verified in the Queensland pneumococcal population. A number of capsule switches had been identified in the earlier studies (Chapter 5). Often identification of capsule

switching is achieved by sequencing the capsule cassette for identification of points of recombination. By using the novel capsule typing method, exact points of recombination cannot be determined, however identification of rearranged genes could be visualised. This study was only able to verify that the serotypes observed were the same as determined by the Quellung reaction. The potential capsule switches observed, including a switch from a serotype 19A to 15C and serotype 1 to 4, has also been observed when combining the novel capsule typing with MLVA4. No evidence of vaccine escape strains has been observed in this study.

6.5 Conclusion and future perspectives

This final study examines the *S. pneumoniae* capsule cassettes, which is a primary target of the current pneumococcal childhood vaccines. The pneumococcal capsule is a major virulence factor, and has been extensively investigated to understand how the pneumococcus has the ability to switch this capsule to evade the current vaccines. The traditional serotyping method for characterising the more than 90 capsule types has a number of limitations, prompting the development of molecular based capsule typing methods. The newly designed capsule typing method has been devised based on the selection of highly discriminatory capsule genes, selected by the bioinformatics program Minimum SNPs. A total of 23 capsule genes could be used to differentiate 93 serotypes, and a minimised 16 target gene was used to characterise the 35 Queensland *S. pneumoniae* serotypes with a 44% success rate. Further optimisation is recommended, although the use of conventional PCR and real-time PCR can be used. Furthermore, the capsule profiles of *S. pneumoniae* isolates that were thought to have been associated with capsule switching (Chapter 5) was verified by using the novel capsule typing method.

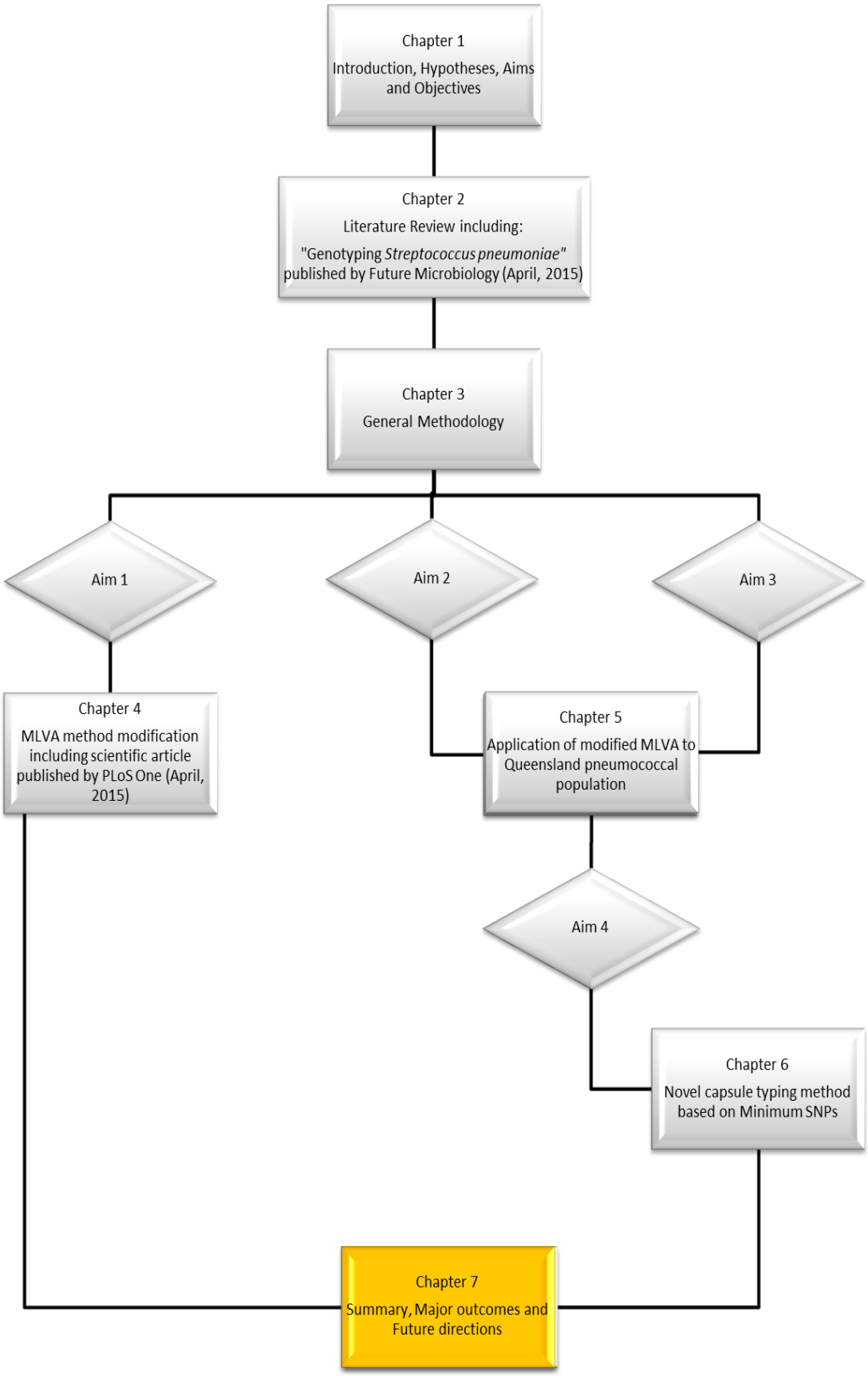
There are a number of future directions for the use of this newly developed capsule typing method. Further optimisation of the method, including the validation of using real-time PCR is required; however results so far are fairly promising. The novel capsular typing method has not been tested with multiple pneumococcal strains with different serotypes. Detection of pneumococcal serotypes in co-colonization (i.e. multiple serotypes in a sample) has been investigated (Saha *et al.,* 2015).

At this point, the novel capsule typing has not been developed into a multiplex; however there is the potential for this. Fluorescent labels can be incorporated to distinguish each primer set within rtPCR. Another study has used fluorescent labels in rtPCR by using FAM, HEX, ROX and CY5 with black hole quencher as reporter dyes

(Pimenta *et al.,* 2013). Alternatively, the use of melt curves can differentiate the different gene products, as we could already see from this study (Bricker, 2011).

It would also be of interest to investigate whether this capsule typing method could be applied directly to clinical samples. Amplification of conserved *wzg* with a TaqMan minor groove binder (MGB) probe was performed directly on pleural fluids taken from patients with parapneumonic empyema caused by *S. pneumoniae* (Tarrago *et al.,* 2008).

CHAPTER 7: SUMMARY, MAJOR OUTCOMES, SIGNIFICANCE AND FUTURE DIRECTIONS


Chapter 1
Introduction, Hypotheses, Aims and Objectives
Chapter 2
Literature Review including:
"Genotyping *Streptococcus pneumoniae*" published by Future Microbiology (April, 2015)
Chapter 3
General Methodology
Aim 1
Aim 2
Aim 3
Chapter 4
MLVA method modification including scientific article published by PLoS One (April, 2015)
Chapter 5
Application of modified MLVA to Queensland pneumococcal population
Aim 4
Chapter 6
Novel capsule typing method based on Minimum SNPs
Chapter 7
Summary, Major outcomes and Future directions

7.1 Summary

S. pneumoniae is undoubtedly an important bacterium to study, especially since the population structure has been changing since the introduction of the childhood pneumococcal vaccines. To monitor these changes, and therefore determine the best advice that health authorities and the public should be given, epidemiology studies are required. Hence this PhD thesis has examined both MLST and MLVA genotyping methods for characterising *S. pneumoniae* in Queensland children, and capsule typing methods for identification of the pneumococcal capsule types that the vaccines target.

To conclude this thesis, the following paragraphs will summarise the four main aims investigated and whether the objectives were met with our major findings. These sections will outline the original hypotheses and whether these were rejected or supported from our results. Following this will be the major outcomes and significance of this study, with concluding remarks of the limitations in this study. A final section is then included on our final recommendations and future directions.

The first aim of this PhD was to modify a MLVA method for genotyping invasive *S. pneumoniae*, with the objective of creating a faster, cheaper, highly discriminatory and technically less-demanding genotyping method compared to the 'gold standard' MLST method. MLST is considered the 'gold standard' genotyping method for studying invasive *S. pneumoniae* however MLVA has also emerged as an alternative genotyping technique as it has higher discriminatory power, is fast and inexpensive (Elberse *et al.,* 2011a; Koeck *et al.,* 2005). Results in this study support this finding; however MLVA would be more suitable for short-term and localised/state-wide epidemiology studies rather than long term and international studies due to the high diversity of the VNTR loci. The modified MLVA4 method maintains a high discriminatory power whilst minimising the number of incomplete genotypes, and is more accurate in representing the Queensland pneumococcal population structure than MLST when observing short-term changes or localised outbreaks. MLST was proven to be less discriminatory (S=0.936) than the MLVA1 protocol (S=0.963) and MLVA4 method (S=0.978). MLVA4 genotyping took less time (3-4 days) compared to MLST (16-20 days) and was inexpensive (~AU$24.00 per isolate compared to ~$AU347.00 per isolate, respectively). Complexes of isolates can be observed in more detail and can glean more information when combined with isolate information such as antibiotic resistance or location of disease (not used in this study). Admittedly, additional markers have increased resolution (ten loci instead of seven or eight) however we have minimised laboratory work to three multiplexes.

The main purpose of modifying MLVA was to eliminate or minimise the number of non-amplified loci detected in Elberse *et al.* (2011a) method, which has been achieved. This study observed that some loci failed to amplify in specific serotypes, such as BOX-06 failing to amplify in 75% of serotype 7F and Spneu19 in several serotype 3 isolates. A number of primers were re-designed in this study (including BOX-10, BOX-12, BOX-13 and Spneu19) in an attempt to successfully amplify the previously failing loci. Improved amplification of these loci was observed after re-designing primers. It was hypothesised that the non-amplification of loci in specific serotypes could provide added discrimination.

Therefore, the use of MLVA rather than MLST for genotyping *S. pneumoniae* should be favoured, particularly due to its low costs, faster time and high discriminatory power. As a result of these advantages, MLVA genotyping is widely used for many organisms. Several microorganisms have limited amount of genetic diversity, therefore VNTR and MLVA methods have been able useful in epidemiology studies because they can allow differentiation of the organisms, including *Mycobacterium tuberculosis* (Spurgiesz *et al.,* 2003), and a number of bioterrorist-organisms such as *Yersinia pestis* (Klevytska *et al.,* 2001), *Bacillus anthracis* (Lista *et al.,* 2006), and *Francisella tularensis* (Farlow *et al.,* 2001; Jernigan *et al.,* 2002). The speed and versatility of MLVA has meant that most bioterrorist-organisms have a MLVA typing scheme (Van Belkum *et al.,* 2006). In addition, VNTR loci have been found in a number of non-bacterial organisms including parasitic protozoa (Wickstead *et al.,* 2003), viruses e.g. Epstein Barr Virus (Xue *et al.,* 2003), and fungus such as *Candida albicans, Aspergillus furnigatus* and *Aspergillus flavus* (Bart-Delabesse *et al.,* 2001; Lott *et al.,* 1999; Wang *et al.,* 2012). Development of MLVA protocols for these non-bacterial organisms is still in the very early stages, but this demonstrates that the application of MLVA is not restricted just to *S. pneumoniae* but to a range of bacterial and non-bacterial organisms. It is also considered favourable for epidemiology studies of outbreaks, and bioterrorism.

The second aim of this study was to determine the population structure of invasive *S. pneumoniae* in Queensland isolated from children 15 years or younger from 2007 to 2012 using MLVA1, MLVA4 and MLST. Our objectives were to observe the differences between MLST, MLVA1 and MLVA4 when interpreting the pneumococcal population structure, and to provide a current study and baseline of the Queensland invasive pneumococcal population.

Variations in interpreting the population structure of *S. pneumoniae* were observed when using the different genotyping protocols. Serotyping alone identifies 35

serotypes but prohibits our understanding of genetic relationships and capsule switching. MLST identifies eleven clonal complexes (n=202) circulating throughout Queensland, whereas MLVA4 identifies eighteen clonal complexes (n=202) and also enables closer examination of eleven MLST singletons, one of which is serotype 7F which is commonly associated with IPD in Australian Aborigines. MLVA4 maintains high congruence with the other MLVA methods and MLST. MLVA4 combined with serotyping, increases our understanding of the Queensland pneumococcal epidemiology. Unfortunately the Queensland MLST results cannot contribute to a national Australian study/database because none currently exists. Few Australian isolates have been submitted to the international MLST database, and we have submitted the MLST results from this study to the MLST database.

Further examination of the MLVA4 pneumococcal population structure reveals diverse genetic relationships within the Queensland population, including between serotypes 18B and 18C (CC7), serotypes 19A, 15C and 15B (CC6), serotype 8 and 11A (CC21), serotype 7F (CC1) and serotype 3 (CC3) which MLST does not reveal. Overlaying patient information with diverse genotypes may reveal further information about outbreaks, patient susceptibility, vaccine immunisation and tracking of disease. The ability to examine CC with higher discrimination using MLVA4 can provide insight into which genes are changing, for example serotype 7F (MLVA CC1 or ST191) largely diversifies due to BOX-10 and serotype 3 diversifies due to Spneu17. It is unknown what the specific functions of these genes are, however it is known that VNTRs and BOX elements play a role in bacterial competence and virulence (Knutsen *et al.,* 2006; Martin *et al.,* 1992). VNTR loci with high diversity (e.g. Spneu17) would allow increased discrimination within localised or short-term studies, whereas VNTR loci with low diversity (e.g. BOX-02 and BOX-11) would allow identification of long term changes.

This epidemiology study so far indicates that the 13vPCV may be having a positive impact by reducing IPD cases in young children. This is one of the first preliminary studies indicating that *S. pneumoniae* may be declining in this group. Unfortunately, IPD is seen to potentially increase in older children. Continued monitoring is essential to determine whether these numbers will decline or not. The significance of these increases in older children may impact vaccine strategy. Overall, there hasn't been a decline in 13vPCV serotypes in children under 15 years old, however it is too soon to detect any significant changes.

In addition we can identify that the four most common serotypes 1, 3, 7F and 19A form the largest clonal complexes circulating in Queensland children from 2007 to 2012. Keeping in mind that not all serotype 19A were genotyped in this study because

they are being examined in another study at QUT, the numbers of serotype 19A are underestimated. Since all of these common serotypes are targeted by 13vPCV, this signifies that IPD cases will hopefully decline in the future with wide implementation of the vaccine. Interestingly, despite the introduction of 13vPCV, a number of 7vPCV-serotypes were still being observed; particularly serotype 19F which was the fifth highest most common serotype in this study. There is 95% vaccine coverage in non-Indigenous citizens and 85% coverage in Indigenous citizens across Australia (Barry *et al.,* 2012). Due to restricted access to patient data for this study, the vaccination history of 7vPCV is unknown for this study; therefore we are unable to determine whether IPD caused by serotype 19F is due to non-vaccination or a vaccine failure. Genotyping also reveals that non-targeted serotypes are genetically related to targeted serotypes e.g. serogroup 15 found in CC5 is genetically related to serotype 19A. This could provide preliminary insight into potential future serotype replacement.

As well as identifying genetic relationships between individual serotypes, higher discriminatory power can impact interpretations of population structures, especially when identifying capsule switches. Already potential capsule switches have been observed between a serotype 19A and 15C in CC6 (MT59; ST411), and a serotype 1 and 4 (MT36; ST306) in CC8 in our Queensland population using MLVA4. Other potential capsule switches were observed using MLVA1 or MLST which may be false positives due to the lower discriminatory power of these genotyping methods, or could actually reflect the limitations of MLVA in detecting all capsule switches. Alternatively, MLVA4 is too discriminatory and even though true capsule switch occurs, MLVA4 identifies two distinct genotypes therefore the assumption is that no capsule switching has occurred. Further investigation is required to determine whether MLVA4 could fail to detect capsule switches or that MLST is detecting false capsule switches. Since MLVA4 is highly discriminatory for short-term studies, it may enable detection of capsule switching earlier than MLST would. Because the serotyping methods may not be discriminatory enough to identify true capsule types, confirmation of capsule switching will be performed by combining a novel capsule typing method that identifies capsule types, with the highly discriminatory MLVA4. The significance of accurately identifying these capsule switches may impact future vaccine development.

These possible limitations of vaccination, i.e. vaccine failure, serotype replacement and capsule switching, could impact the use of vaccines for combating disease in the future. As demonstrated by the pneumococcal vaccine and how this target organism has almost countermanded the effects, this could inform that careful consideration is required when developing future vaccines. A number of other

microorganisms have polysaccharide capsules similar to the pneumococcus, including Gram-negative bacteria such as *Escherichia coli* (Bayer & Thurow, 1997; Sarkar *et al.,* 2014), *Neisseria meningitidis* (Harrison *et al.,* 2013), *Klebsiella pneumoniae* (Domenico *et al.,* 1994), *Haemophilus influenzae* (Kroll *et al.,* 1990), *Pseudomonas aeruginosa* (Evans & Linker, 1973), *Salmonella enterica* (Gibson *et al.,* 2006), Gram-positive bacteria such as *Bacillus megaterium* (Baumann-Grace & Tomscik, 1957), *Streptococcus pyogenes* (Obzanska *et al.,* 2011), *Streptococcus agalactiae* (Bellais *et al.,* 2012), *Staphylococcus epidermidis* (Ziebuhr *et al.,* 1999), *Staphylococcus aureus* (O'Riordan & Lee, 2004), and even fungus such as *Cryptococcus neoformans* (O'Meara & Alspaugh, 2012). As a result, vaccines have or are being developed to target the polysaccharide capsules of these microorganisms, however as demonstrated by the effects of the pneumococcal vaccines on *S. pneumoniae* a number of concerns remain. Evidence of capsule switching have been observed in other microorganisms such as *Streptococcus agalactiae* (Bellais *et al.,* 2012) and *Neisseria meningitidis* (Swartley *et al.,* 1997), and the different types of capsules make vaccine design difficult, for example *E. coli* has 167 different O serogroups and more than 80 K polysaccharide antigens (Brumbaugh & Mobley, 2012), and more than 20 different serotypes for *Pseudomonas aeruginosa* (Sharma *et al.,* 2013). Certainly in the case of the pneumococcus, developing a vaccine that only targets a smaller set number of serotypes has presented some limitations when combating pneumococcal disease. The vaccine has systematically reduced the incidence of IPD among the target-vaccinated group in several countries. However, the vaccine has also become a selection pressure against the pneumococcal population, causing it to shift from a targeted-serotype population to a non-targeted serotype population. On top of this, capsule switching in the pneumococcal populations have been observed worldwide, leading to vaccine escape strains that cause increase number of pneumococcal disease. It appears that any approach to designing a vaccine, not even just for *S. pneumoniae* but for any organism, needs to achieve complete eradication by allowing minimal opportunities to open a niche for the same organism to adapt to.

Following from the previous aims, Aim 3 was to determine whether the genetic population of non-vaccine (non-13vPCV) serotypes was changing in the Queensland pneumococcal population. By comparing eBurst dendograms of 2007-2009 to 2010-2012 demonstrates that the population structure appears to be less clonal in later years, supporting our original hypothesis that the population structure would be changing. This is significant as it indicates that strains are diversifying and increasing in numbers. The objective of the vaccine is to reverse this trend. As it is still too early to

determine the significant changes after the introduction of the 13vPCV, these results are only preliminary observations.

Detecting relationships between serotypes accurately and quickly may have an impact on the selection of serotypes for future vaccine strategies. Ideally vaccine strategies should benefit the whole of a nation, however differences in state pneumococcal populations have prompted localised strategies, for example the 10vPCV was only introduced into Northern Territory since the additional serotypes were not common elsewhere in Australia. If a vaccine targets only one serotype in a CC and not the other, this may provide pressure for capsule switching or serotype replacement. The 13vPCV targets serotype 19A, therefore there might be a selective pressure for the serotype 19A to switch capsules to a 15C or 15B (CC6), which are known to be genetically related. Serotype 15B and 15C are not included in any current childhood vaccines, or the in-trial 15vPCV which targets the extra serotypes 22F and 33F, therefore future surveillance is recommended.

Conversely, a number of clonal complexes have emerged, indicating the possible beginnings of serotype replacement, especially those complexes containing serotypes not targeted by the 13vPCV. It will therefore be important to monitor the potential changes of CC9 (serotype 15C), CC10 (serotype 15B/15C), CC32 (serotype 33F), CC19 (serotype 22F and 33F) and CC24 (serotype 22F). Continued surveillance for changes to these clonal complexes, especially CC19 and CC24, will assist in vaccine health policy decision making and the impact of introducing the 15vPCV (targets serotype 22F and 33F).

Despite the emergence of clonal complexes, serotype replacement has not been observed since the introduction of the 13vPCV in 2011 in this study. However there have been increases of serotype 6C and 23B, both not targeted by the 13vPCV. Altogether, serotype 6C, 15C and 33F (9% of isolates) could increase by serotype replacement as they are not targeted by the 13vPCV. It is recommended that surveillance by way of MLVA continue to enable reporting of future serotype replacement. Cooper *et al.* (2011) has demonstrated that the 13vPCV may have cross-protection against 6C as there was a 96% OPA (238 opsonophagocytic assay) response. In South Australia there have also been reports of increases in non-13vPCV serotypes (Johnson *et al.,* 2012).

The final aim of this study builds on the genotyping study in Aims 1, 2 and 3 by combining a novel capsule typing method developed in this study with MLVA4 to determine capsule types and verify capsule switches. Examination of 93 *S. pneumoniae* capsular cassettes available in the NCBI database was performed so that a 'mother

template' consisting of all the possible capsular genes could be designed. Using the Minimum SNPs bioinformatics program, a binary marker system of T = present and A = absent was used to create a pseudoDNA sequence for each of the 93 capsule types based from the 'mother template'. The Minimum SNPs program then analysed the combination of binary markers and identified a minimum set of targets (capsule genes) to characterise each capsule type. The minimum number of targets identified in this study to characterise Queensland invasive *S. pneumoniae* was 16 capsule genes. These selected capsule genes were applied to clinical strains of *S. pneumoniae* isolated from children less than 15 years old during 2007 to 2012. Conventional PCR and real-time PCR were both used to demonstrate that the capsule typing method could be applicable to both techniques.

The newly developed novel capsule typing method was applied to 48 isolates, with a 44% successful typeability when using conventional PCR. Closer examination of the targeted capsule genes identify that most had good typeability, while others will require further optimisation. As well as standard capsule typing, the novel capsule typing was combined with MLVA4 to verify capsule switching events previously observed in the Queensland pneumococcal population structure from Aim 3 (Chapter 5). Verification of the capsule switch between a serotype 1 and 4, and serotypes 19A and 15C was observed. Both serotype 1 and 4 appeared to have a different capsule type profile although they both had the same MLVA4 genotype. The application of sequential PCR has also been demonstrated for the capsule typing method. A set of five genes (*wzg, rmlB, glf, wchA* and *wchF*) can differentiate the majority of serotypes in complexes, of which only a couple of other genes or SNP-specific primers that can be utilised to further differentiate the complexes.

Future directions of the novel capsule typing include the application to co-colonisations of *S. pneumoniae* serotype, direct typing of clinical specimens without the need for culture, and the identification of non-capsular *S. pneumoniae* that do not have an expressed capsule. The wider application of molecular capsule typing to other microorganisms could be performed, particularly those mentioned before that also carry a polysaccharide capsule. The Minimum SNPs bioinformatics program can be similarly applied to other microorganisms to determine the minimum number of capsule genes that would provide the highest discrimination between all capsule types and sub-types.

7.2 Major Outcomes and Significance of this study

1. Modification of MLVA genotyping method for *S. pneumoniae* offers a rapid, cost-effective and highly discriminatory alternative to MLST. MLVA4 also decreases the number of incomplete genetic profiles that previous published MLVA methods obtained. Any non-amplifications were observed to be serotype specific and possibly due to absence of the locus in the pneumococcal genome. It is possible to standardise this method between laboratories which will enable the sharing of data.

2. The modified MLVA4 genotyping method can be used for epidemiology studies of invasive *S. pneumoniae* population structure. It could also be applicable to carriage isolates of *S. pneumoniae*. Importantly, this MLVA4 protocol has enabled further subdivision of important pneumococcal serotypes and sequence types causing disease in Queensland children, including serotype 7F and serotype 19A.

3. A number of new sequence types (ST), namely ST8969, ST9082, and ST9660 – ST9673 for *S. pneumoniae* were characterised from the Queensland population. Full MLST sequencing was carried out and they are now listed on the MLST database (http://pubmlst.org/spneumoniae/).

4. A novel capsule typing method was developed to identify the pneumococcal serotypes. This new technique was based on a novel approach by using Minimum SNPs program to identify a minimum set of highly discriminatory genes to characterise each known serotype of *S. pneumoniae*. Real-time PCR was used to determine the absence or presence of each selected gene. The application of Minimum SNPs for molecular capsule typing of other microorganisms based on capsular genes is also feasible.

5. The novel capsule typing was applied to Queensland isolates of *S. pneumoniae*. Serotypes could be determined using this novel method. This new technique enables a rapid and cost-effective alternative to traditional serotyping methods, and has the ability to capsule type all known serotypes whereas other PCR based molecular typing methods do not.

6. The combination of the novel capsule typing with the modified MLVA4 genotyping can be used in epidemiology studies, particularly to identify capsule switching. Several capsule switching events were identified in this study, namely serotype 19A to 15C, and serotype 1 to 4.

7. This research increases our knowledge of the pneumococcal population structure of IPD cases in children 15 years or younger in Queensland, and emphasises the need for a universal genotyping method and database.

7.3 Future directions

1. After the introduction of a vaccine, it is desirable to survey a population for at least five years post introduction; therefore MLVA4 genotyping could be performed for 2013-2016 to determine the population structure and determine shifts in the population structure post-13vPCV.
2. Ethical clearance would be desirable so that further examination of the associated patient data with the pneumococcal population can be performed. Information such as location, vaccination history, age, gender, etc. could have association with the rates of IPD. Furthermore examination of antibiotic resistance and MDR strains particularly in non-13vPCV serotypes can be examined.
3. Due to the lack of a national genetic database in Australia (either MLST or MLVA), further collaboration to investigate or establish a universal MLVA genotyping method and database for genotyping *S. pneumoniae* is recommended.
4. Sequencing of unusual VNTR fragment sizes in serotype 33F, 35F and 7F can be performed. This could determine whether the locus is missing or contains large insertion sequence (IS) fragments. Study of VNTR genes will be important particularly because of their association with virulence genes. Additionally, experimentation of the VNTR/BOX locus used in this study and its association with downstream loci can be performed to determine whether they play regulatory roles in virulence.
5. A novel capsule typing method has been developed in this study but has only been applied to 35 serotypes available in Queensland, and validated using conventional PCR. It would be of interest to further optimise this method by validating it using rtPCR, transforming it into multiplex reactions (and potentially sequentially like Pai *et al.* 2006) and apply the method directly to clinical samples without the need for culture, and non-capsulated pneumococci. Application of rtPCR was attempted in this study, but due to difficulties in amplifying capsule genes and time constraints, validation of the method was not performed. By multiplexing the reactions, this would provide a faster and potentially cheaper method. The ability to capsule type directly from clinical

samples (e.g. blood, sputum, etc.) without the need to culture can decrease time and cost of experimentation and furthermore being able to detect non-capsulated *S. pneumoniae* is equally important as these strains are difficult to detect and yet also cause invasive disease. Furthermore, the capsule typing method needs to be tested against other Streptococcal species, particularly *S. mitis* which has been shown to test positive.

6. Verification of whether the novel capsule typing can be confidently used to determine capsule switching needs to be compared to whole genome sequencing or capsule sequencing.
7. Exploration into particular serotypes, for example serotype 1 which has a potential capsule switch from serotype 4, or serogroup 6 which now has 8 distinct capsular polysaccharides (A, B, C, D, E, F, G, H) could be performed.

REFERENCES

Aanensen, D. M., & Spratt, B. G. (2005). The multilocus sequence typing network: mlst.net. *Nucleic Acids Research, 33*, 728-733. doi: 10.1093/nar/gki415.

Ahn, J. G., Choi, S. Y., Kim, D. S., & Kim, K. H. (2012). Enhanced detection and serotyping of *Streptococcus pneumoniae* using multiplex polymerase chain reaction. *Korean Journal of Pediatrics, 55*(11), 424-429.

Ansaldi, F., Canepa, P., de Florentiis, D., Bandettini, R., Durando, P., & Icardi, G. (2011). Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine. *Clinical and Vaccine Immunology, 18*(2), 343-345. doi: 10.1128/cvi.00383-10.

Antonio, M., Hakeem, I., Sankareh, K., Cheung, Y. B., & Adegbola, R. A. (2009). Evaluation of sequential multiplex PCR for direct detection of multiple serotypes of *Streptococcus pneumoniae* from nasopharyngeal secretions. *Journal of Medical Microbiology, 58*(3), 296-302.

Austrian, R. (1977). Prevention of pneumococcal infection by immunization with capsular polysaccharides of *Streptococcus pneumoniae*: current status of polyvalent vaccines. *Journal of Infectious Diseases, 136*(Suppl.), 38-42.

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus types III. *The Journal of Experimental Medicine, 79*(2), 137-158.

Azzari, C., Moriondo, M., Indolfi, G., Cortimiglia, M., Canessa, C., Becciolini, L., . . . Resti, M. (2010). Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. *PLoS One, 5*(2). doi:10.1371/journal.pone.0009282

Barry, C., Krause, V., Cook, H., & Menzies, R. (2012). Invasive pneumococcal disease in Australia 2007 and 2008. *Communicable Diseases Intelligence, 36*(2).

Bart-Delabesse, E., Sarfati, J., Debeaupuis, J. P., Van Leeuwen, W., Van Belkum, A., Bretagne, S., & Latge, J. P. (2001). Comparison of restriction fragment length polymorphism, microsatellite length polymorphism, and random amplification of polymorphic DNA analyses for fingerprinting *Aspergillus fumigatus* isolates. *Journal of Clinical Microbiology, 39*(7), 2683-2686. doi: 10.1128/JCM.39.7.2683-2686.2001.

Baumann-Grace, J. B., & Tomcsik, J. (1957). The surface structure and serological typing of *Bacillus megaterium*. *Journal of General Microbiology, 17*(1), 227-237. doi: 10.1099/00221287-17-1-227.

Bayer, M. E., & Thurow, H. (1977). Polysaccharide capsule of *Escherichia coli*: microscope study of its size, structure, and sites of synthesis. *Journal of Bacteriology, 130*(2), 911-936.

Beall, B., McEllistrem, M. C., Gertz, J. R. E., Wedel, S., Boxrud, D. J., Gonzalez, A. L., . . . Active Bacterial Core Surveillance. (2006). Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *Journal of Clinical Microbiology, 44*(3), 999-1017.

Bellais, S., Six, A., Fouet, A., Longo, M., Dmytruk, N., Glaser, P., . . . Poyart, C. (2012). Capsular switching in group B streptococcus CC17 hypervirulent clone: A future challenge for polysaccharide vaccine development. *Journal of Infectious Diseases, 206*(11), 1745-1752. doi: 10.1093/infdis/jis605.

Benson, G. (1999). Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Research, 27*(2), 573-580.

Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., . . . Spratt, B. G. (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genetics, 2*(3), e31. doi:10.1371/journal.pgen.0020031.

Berglund, J., Vink, P., Fernanda Tavares Da, S., Lestrate, P., & Boutriau, D. (2014). Safety, Immunogenicity, and Antibody Persistence following an Investigational *Streptococcus pneumoniae* and *Haemophilus influenzae* Triple-Protein Vaccine in a Phase 1 Randomized Controlled Study in Healthy Adults. *Clinical and Vaccine Immunology, 21*(1), 56-65.

Bratcher, P. E., Kim, K., Kang, J. H., Hong, J. Y., & Nahm, M. H. (2010). Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. *Microbiology, 156*(2), 555-560.

Bratcher, P. E., Park, I. H., Oliver, M. B., Hortal, M., Camilli, R., Hollingshead, S. K., . . . Nahm, M. H. (2011). Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. *Microbiology, 157*(1), 189-198.

Bricker, B. (2011). Past, present and future of molecular technology applications for the epidemiology of bacterial diseases. *Journal of Analytical and Bioanalytical Techniques, S10:001*. doi:10.4172/2155-9872.S10-001.

Brito, D. A., Ramirez, M., & de Lencastre, H. (2003). Serotyping *Streptococcus pneumoniae* by multiplex PCR. *Journal of Clinical Microbiology, 41*(6), 2378-2384.

Brueggemann, A. B., & Spratt, B. G. (2003). Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. *Journal of Clinical Microbiology, 41*(11), 4966-4970.

Brueggemann, A. B., Pai, R., Crook, D. W., & Beall, B. (2007). Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathogens, 3*(11), e168. doi:10.1371/journal.ppat.0030168.

Brugger, S. D., Hathaway, L. J., & Muhlemann, K. (2009). Detection of *Streptococcus pneumoniae* strain cocolonization in the nasopharynx. *Journal of Clinical Microbiology, 47*(6), 1750-1756. doi: 10.1128/JCM.01877-08.

Brumbaugh, A. R., & Mobley, H. L. T. (2012). Preventing urinary tract infection: progress toward an effective *Escherichia coli* vaccine. *Expert Review of Vaccines, 11*(6), 663-676. doi: 10.1586/ERV.12.36.

Calix, J. J., & Nahm, M. H. (2010). A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *The Journal of Infectious Diseases, 202*(1), 29-38.

Calix, J. J., Porambo, R. J., Brady, A. M., Larson, T. R., Yother, J., Abeygunwardana, C., & Nahm, M. H. (2012). Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *Journal of Biological Chemistry, 287*(33), 27885.

Camara, M., Mitchell, T. J., Andrew, P. W., & Boulnois, G. J. (1991). *Streptococcus pneumoniae* produces at least two distinct enzymes with neuraminidase activity: cloning and expression of a second neuraminidase gene in *Escherichia coli*. *Infection and Immunity, 59*(8), 2856-2858.

Carvalho, M. d. G., Hollingshead, S. K., Pilishvili, T., Whitney, C. G., Nahm, M. H., Beall, B. W., . . . Park, I. H. (2009). PCR-based quantitation and clonal diversity of the current prevalent invasive serogroup 6 pneumococcal serotype, 6C, in the United States in 1999 and 2006 to 2007. *Journal of Clinical Microbiology, 47*(3), 554-559.

Carvalho, M. D. G., Pimenta, F. C., Moura, I., Roundtree, A., Gertz, R. E., Li, Z. Y., . . . Beall, B. (2013). Non-pneumococcal mitis-group streptococci confound detection of pneumococcal capsular serotype-specific loci in upper respiratory tract. *PeerJ, 1*(1), e97. doi:10.7717/peerj.97.

Centers for Disease Control Prevention. (2010). Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children – Advisory Committee on Immunization Practices (ACIP), 2010. *Morbidity and Mortality Weekly Report, 59*(9), 258-261.

Centers for Disease Control Prevention. (2015). Invasive pneumococcal disease (IPD) (*Streptococcus pneumoniae*, invasive disease) (*Streptococcus pneumoniae*): 2010 Case definition. Retrieved 28 July, 2015, from http://wwwn.cdc.gov/NNDSS/script/casedef.aspx?CondYrID=736&DatePub=1/1/2010%2012:00:00%20AM.

Chiavolini, D., Memmi, G., Maggi, T., Iannelli, F., Pozzi, G., & Oggioni, M. R. (2003). The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. *BMC Microbiology, 3*(1), 14-14.

Claverys, J. P., & Martin, B. (1998). Competence regulons, genomics and streptococci. *Molecular Microbiology, 29*(4), 1126-1127.

Cohen, R., Levy, C., Bingen, E., Koskas, M., Nave, I., & Varon, E. (2012). Impact of 13-valent pneumococcal conjugate vaccine on pneumococcal nasopharyngeal carriage in children with acute otitis media. *Pediatric Infectious Disease Journal, 31*, 297-301.

Cohen, R., Levy, C., Bonnet, E., Thollot, F., Boucherat, M., Fritzell, B., . . . Varon, E. (2011). Risk factors for serotype 19A carriage after introduction of 7-valent pneumococcal vaccination. *BMC Infectious Diseases, 11*(1), 95-95.

Collins, D., Hoskins, A., Bowman, J., Jones, J., Stemberger, N., Richmond, P., . . . Lehmann, D. (2013). High nasopharyngeal carriage of non-vaccine serotypes in Western Australian Aboriginal people following 10 years of pneumococcal conjugate vaccination: e82280. *PLoS One, 8*(12). doi:10.1371/journal.pone.0082280.

Cooper, D., Yu, X., Sidhu, M., Nahm, M. H., Fernsten, P., & Jansen, K. U. (2011). The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to *Streptococcus pneumoniae* serotypes 6C and 7A. *Vaccine, 29*(41), 7207-7211.

Cornick, J. E., & Bentley, S. D. (2012). *Streptococcus pneumoniae*: The evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes and Infection, 14*(7-8), 573-583.

Coskun-Ari, F. F., Guldemir, D., & Durmaz, R. (2012). One-step multiplex PCR assay for detecting *Streptococcus pneumoniae* serogroups/types covered by 13-valent pneumococcal conjugate vaccine (PCV13). *PLoS One, 7*(12). doi:10.1371/journal.pone.0050406.

Croucher, N. J., Harris, S. R., Barquist, L., Parkhill, J., & Bentley, S. D. (2012). A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathogens, 8*(6), e1002745. doi:10.1371/journal.ppat.1002745.

Croucher, N. J., Vernikos, G. S., Parkhill, J., & Bentley, S. D. (2011). Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics, 12*(1), 120-120.

Davis, S. M., Deloria-Knoll, M., Kassa, H. T., & O'Brien, K. L. (2013). Impact of pneumococcal conjugate vaccines on nasopharyngeal carriage and invasive disease among unvaccinated people: review of evidence on indirect effects. *Vaccine, 32*(1), 133-145.

Denoël, P., Philipp, M. T., Doyle, L., Martin, D., Carletti, G., & Poolman, J. T. (2011). A protein-based pneumococcal vaccine protects rhesus macaques from pneumonia after experimental infection with *Streptococcus pneumoniae*. *Vaccine, 29*(33), 5495-5501.

Dias, C. A., Teixeira, L. M., Carvalho, M. d. G., & Beall, B. (2007). Sequential multiplex PCR for determining capsular serotypes of pneumococci recovered from Brazilian children. *Journal of Medical Microbiology, 56*(9), 1185-1188.

Domenico, P., Salo, R. J., Cross, A. S., & Cunha, B. A. (1994). Polysaccharide capsule-mediated resistance to opsonophagocytosis in *Klebsiella pneumoniae*. *Infection and Immunity, 62*(10), 4495-4499.

Duggan, S. T. (2010). Pneumococcal polysaccharide conjugate vaccine (13-valent, adsorbed) [Prevenar 13®]. *Drugs, 70*(15), 1973-1986.

Elberse, K. E. M., van der Heide, H. G. J., Witteveen, S., van de Pol, I., Schot, C. S., van der Ende, A., . . . Schouls, L. M. (2012). Changes in the composition of the pneumococcal population and in IPD incidence in The Netherlands after the implementation of the 7-valent pneumococcal conjugate vaccine. *Vaccine, 30*(52), 7644-7651.

Elberse, K. E. M., Nunes, S., Sa-Leao, R., van der Heide, H. G. J., & Schouls, L. M. (2011). Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. *PLoS One, 6*(5), e19668. doi:10.1371/journal.pone.0019668.

Elberse, K. E. M., van de Pol, I., Witteveen, S., van der Heide, H. G. J., Schot, C. S., van Dijk, A., . . . Schouls, L. M. (2011). Population structure of invasive *Streptococcus pneumoniae* in The Netherlands in the pre-vaccination era assessed by MLVA and capsular sequence typing. *PLoS One, 6*(5), e20390. doi:10.1371/journal.pone.0020390.

Elberse, K., Witteveen, S., van der Heide, H., van de Pol, I., Schot, C., van der Ende, A., . . . Schouls, L. (2011). Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. *PLoS One, 6*(9), e25018. doi:10.1371/journal.pone.0025018.

Enright, M. C., & Spratt, B. G. (1998). A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology, 144*(11), 3049-3060.

Evans, L. R., & Linker, A. (1973). Production and characterization of the slime polysaccharide of *Pseudomonas aeruginosa*. *Journal of Bacteriology, 116*(2), 915-924.

Fagan, R. L., Hanna, J. N., Messer, R. D., Brookes, D. L., & Murphy, D. M. (2001). The epidemiology of invasive pneumococcal disease in children in far north Queensland. *Journal of Paediatrics and Child Health, 37*(6), 571-575.

Farlow, J., Smith, K. L., Wong, J., Abrams, M., Lytle, M., & Keim, P. (2001). *Francisella tularensis* strain typing using Multiple-Locus, Variable-Number Tandem Repeat Analysis. *Journal of Clinical Microbiology, 39*(9), 3186-3192. doi: 10.1128/JCM.39.9.3186-3192.2001.

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., & Carrico, J. A. (2012). PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics, 13*(1), 87.

Gibson, D. L., White, A. P., Snyder, S. D., Martin, S., Heiss, C., Azadi, P., . . . Kay, W. W. (2006). *Salmonella* produces an O-antigen capsule regulated by AgfD and important for environmental persistence. *Journal of Bacteriology, 188*(22), 7722-7730. doi: 10.1128/JB.00809-06.

Giele, C., Moore, H., Bayley, K., Harrison, C., Murphy, D., Rooney, K., . . . Lehmann, D. (2007). Has the seven-valent pneumococcal conjugate vaccine had an impact on invasive pneumococcal disease in Western Australia? *Vaccine, 25*(13), 2379-2384.

Goldblatt, D., Southern, J., Ashton, L., Richmond, P., Burbidge, P., Tasevska, J., . . . Miller, E. (2006). Immunogenicity and boosting after a reduced number of doses of a pneumococcal conjugate vaccine in infants and toddlers. *The Pediatric Infectious Disease Journal, 25*(4), 312-319.

Golubchik, T., Brueggemann, A. B., Street, T., Gertz, J. R. E., Spencer, C. C. A., Ho, T., . . . Bowden, R. (2012). Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nature Genetics, 44*(3), 352-356.

Gounder, P. P., Bruce, M. G., Bruden, D., Singleton, R., Rudolph, K., Hurlburt, D. A., . . . Wenger, J. D. (2014). Effect of the 13-valent pneumococcal conjugate vaccine on nasopharyngeal colonization by *Streptococcus pneumoniae* - Alaska, 2008 - 2012. *Journal of Infectious Diseases, 209*(8), 1251-1258.

Gratten, M., Morey, F., Dixon, J., Torzillo, P., & Erlich, J. (1995). Invasive type 12F pneumococcal disease in central Australia. *Communicable Diseases Intelligence, 19*(19), 470-473.

Gratten, M., Nimmo, G., Carlisle, J., Schooneveldt, J., Seneviratne, E., Kelly, R., . . . Fenton, J. (1997). Emergence of further serotypes of multiple drug-resistant *Streptococcus pneumoniae* in Queensland. *Communicable Diseases Intelligence, 21*(10), 133-136.

Griffith, F. (1928). The significance of pneumococcal types. *Journal of Hygiene, 27*(2), 113-159.

Guevara, M., Barricarte, A., Gil-Setas, A., García-Irure, J. J., Beristain, X., Torroba, L., . . . Castilla, J. (2009). Changing epidemiology of invasive pneumococcal disease following increased coverage with the heptavalent conjugate vaccine in Navarre, Spain. *Clinical Microbiology and Infection, 15*(11), 1013-1013.

Hall, L. M. C. (1998). Application of molecular typing to the epidemiology of *Streptococcus pneumoniae*. *Journal of Clinical Pathology, 51*, 270-274.

Hanage, W. P., Kaijalainen, T. H., Syrjanen, R. K., Auranen, K., Leinonen, M., Makela, P. H., & Spratt, B. G. (2005). Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infection and Immunity, 73*(1), 431-435.

Hanna, J. N., Gratten, M., Tiley, S. M., Brookes, D. L., & Bapty, G. (1997). Pneumococcal vaccination: an important strategy to prevent pneumonia in Aboriginal and Torres Strait Island adults. *Australian and New Zealand Journal of Public Health, 21*(3), 281-285.

Hanna, J. N., Humphreys, J. L., & Murphy, D. M. (2006). Invasive pneumococcal disease in Indigenous people in north Queensland, 1999-2004. *Medical Journal of Australia, 184*(3), 118-121.

Hanna, J. N., Humphreys, J. L., & Murphy, D. M. (2008). Invasive pneumococcal disease in Indigenous people in north Queensland: an update, 2005-2007. *Medical Journal of Australia, 189*(1), 43-46.

Hanna, J. N., Humphreys, J. L., Murphy, D. M., & Smith, H. V. (2010). Invasive pneumococcal disease in non-Indigenous people in north Queensland, 2001-2009. *Medical Journal of Australia, 193*(7), 392-396.

Hansman, D., & Bullen, M. M. (1967). A resistant pneumococcus. *Lancet, 2*, 264-265.

Harrison, O. B., Claus, H., Jiang, Y., Bennett, J. S., Bratcher, H. B., Jolley, K. A., . . . Maiden, M. C. J. (2013). Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerging Infectious Diseases, 19*(4), 566-573. doi: 10.3201/eid1904.111799.

Hausdorff, W. P., Bryant, J., Paradiso, P. R., & Siber, G. R. (2000). Which Pneumococcal Serogroups Cause the Most Invasive Disease: Implications for Conjugate Vaccine Formulation and Use, Part I. *Clinical Infectious Diseases, 30*(1), 100-121.

Hausdorff, W. P., Bryant, J., Kloek, C., Paradiso, P. R., & Siber, G. R. (2000). The Contribution of Specific Pneumococcal Serogroups to Different Disease Manifestations: Implications for Conjugate Vaccine Formulation and Use, Part II. *Clinical Infectious Diseases, 30*(1), 122-140.

Hermans, P. W., Sluijter, M., Hoogenboezem, T., Heersma, H., van Belkum, A., & de Groot, R. (1995). Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. *Journal of Clinical Microbiology, 33*(6), 1606-1612.

Hicks, L. A., Harrison, L. H., Flannery, B., Hadler, J. L., Schaffner, W., Craig, A. S., . . . Active Bacterial Core Surveillance Program of the Emerging Infections Program. (2007). Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. *The Journal of Infectious Diseases, 196*(9), 1346-1354.

Hills, S. L., Hanna, J. N., & Murphy, D. (2002). Invasive pneumococcal disease in north Queensland, 2001. *Communicable Diseases Intelligence, 26*(4), 520-524.

Hoang, V.-T., Gaëlle, C., Katia, H., Brigitte, F., Harriet, C., Annick, C., . . . Christine, P. (2007). Multiple-Locus Variable-Number Tandem-Repeat Analysis for longitudinal survey of sources of *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *Journal of Clinical Microbiology, 45*(10), 3175-3183.

Holleley, C., & Geerts, P. G. (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques, 46*(7), 511-517.

Hoskins, J., Fuller, W., Geringer, C., Gilmour, R., Glass, J. S., Khoja, H., . . . Fu, D. J. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *Journal of Bacteriology, 183*(19), 5709-5717.

Hui, F. M., & Morrison, D. A. (1991). Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows *comA*, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *Journal of Bacteriology, 173*(1), 372-381.

Hunter, P. R., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology, 26*(11), 2465-2466.

Hyams, C., Trzcinski, K., Camberlein, E., Weinberger, D. M., Chimalapati, S., Noursadeghi, M., . . . Brown, J. S. (2013). *Streptococcus pneumoniae* capsular serotype invasiveness correlates with the degree of Factor H binding and opsonization with C3b/iC3b. *Infection and Immunity, 81*(1), 354-363.

Hyytiä-Trees, E. K., Cooper, K., Ribot, E. M., & Gerner-Smidt, P. (2007). Recent developments and future prospects in subtyping of foodborne bacterial pathogens. *Future Microbiology, 2*(2), 175-185.

Ichikawa, K., Yagi, T., Inagaki, T., Moriyama, M., Nakagawa, T., Uchiya, K.-i., . . . Ogawa, K. (2010). Molecular typing of *Mycobacterium intracellulare* using multilocus variable-number of

tandem-repeat analysis: identification of loci and analysis of clinical isolates. *Microbiology, 156*(2), 496-504.

Iraurgui, P., Torres, M. J., Gandia, A., Vazquez, I., Cabrera, E. G., Obando, I., . . . Aznar, J. (2010). Modified sequential multiplex PCR for determining capsular serotypes of invasive pneumococci recovered from Seville. *Clinical Microbiology and Infection, 16*(9), 1504-1507.

Isaacman, D. J., Fletcher, M. A., Fritzell, B., Ciuryla, V., & Schranz, J. (2007). Indirect effects associated with widespread vaccination of infants with heptavalent pneumococcal conjugate vaccine (PCV7; Prevnar). *Vaccine, 25*(13), 2420-2427.

Iyer, R., Baliga, N. S., & Camilli, A. (2005). Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in *Streptococcus pneumoniae*. *Journal of Bacteriology, 187*(24), 8340-8349.

Jacobs, M. R., Good, C. E., Bajaksouzian, S., & Windau, A. R. (2008). Emergence of *Streptococcus pneumoniae* serotypes 19A, 6C, and 22F and serogroup 15 in Cleveland, Ohio, in relation to introduction of the protein-conjugated pneumococcal vaccine. *Clinical Infectious Diseases, 47*(11), 1388.

Jacobs, M. R., Koornhof, H. J., Robins-Browne, R. M., Stevenson, C. M., Vermaak, Z. A., Freiman, I., . . . Austrian, R. (1978). Emergence of multiply resistant pneumococci. *The New England Journal of Medicine, 299*(14), 735.

Jacups, S. P., & Cheng, A. (2011). The epidemiology of community acquired bacteremic pneumonia, due to *Streptococcus pneumoniae*, in the top end of the Northern Territory, Australia-over 22 years. *Vaccine, 29*(33), 5386-5392.

Jefferies, J., Clarke, S. C., Diggle, M. A., Smith, A., Dowson, C., & Mitchell, T. (2003). Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. *Molecular Biotechnology, 24*(3), 303-307.

Jefferies, J. M. C., Smith, A., Clarke, S. C., Dowson, C., & Mitchell, T. J. (2004). Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. *Journal of Clinical Microbiology, 42*(12), 5681-5688.

Jernigan, D. B., Raghunathan, P. L., Bell, B. P., Brechner, R., Bresnitz, E. A., Butler, J. C., . . . National Anthrax Epidemiologic Investigation. (2002). Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiologic findings. *Emerging Infectious Diseases, 8*(10), 1019-1028. doi: 10.3201/eid0810.020353.

Jiang, S.-M., Wang, L., & Reeves, P. R. (2001). Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene complexs. *Infection and Immunity, 69*(3), 1244-1255.

Jin, P., Xiao, M., Kong, F., Oftadeh, S., Zhou, F., Liu, C., & Gilbert, G. L. (2009). Simple, accurate, serotype-specific PCR assay to differentiate *Streptococcus pneumoniae* serotypes 6A, 6B, and 6C. *Journal of Clinical Microbiology, 47*(8), 2470-2474.

Johnson, D. R., D'Onise, K., Holland, R. A., Raupach, J. C. A., & Koehler, A. P. (2012). Pneumococcal disease in South Australia: vaccine success but no time for complacency. *Vaccine, 30*(12), 2206-2211.

Johnson, H. L., Deloria-Knoll, M., Levine, O. S., Stoszek, S. K., Freimanis Hance, L., Reithinger, R., . . . O'Brien, K. L. (2010). Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Medicine, 7*(10), e1000348. doi:10.1371/journal.pmed.1000348.

Johnston, C., Campo, N., Bergé, M. J., Polard, P., & Claverys, J.-P. (2014). *Streptococcus pneumoniae*, le transformiste. *Trends in Microbiology, 22*(3), 113-119.

Jones, C., & Lemercinier, X. (2005). Full NMR assignment and revised structure for the capsular polysaccharide from *Streptococcus pneumoniae* type 15B. *Carbohydrate Research, 340*(3), 403-409.

Jorgensen, G., Singleton, R., Butler, J., & Cottle, T. (2005). Outbreak of invasive pneumococcal disease - Alaska, 2003-2004. *Morbidity and Mortality Weekly Report, 54*(3), 72-75.

Jourdain, S., Drèze, P.-A., Vandeven, J., Verhaegen, J., Van Melderen, L., & Smeesters, P. R. (2011). Sequential multiplex PCR assay for determining capsular serotypes of colonizing *S. pneumoniae*. *BMC Infectious Diseases, 11*(1), 100-100.

Kadioglu, A., Weiser, J. N., Paton, J. C., & Andrew, P. W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Microbiology, 6*(4), 288-301.

Kang, H. P., & Dunne, W. M. (2003). Stability of repetitive-sequence PCR patterns with respect to culture age and subculture frequency. *Journal of Clinical Microbiology, 41*(6), 2694-2696.

Kaplan, S. L., Barson, W. J., Lin, P. L., Romero, J. R., Bradley, J. S., Tan, T. Q., . . . Mason, J. E. O. (2013). Early trends for invasive pneumococcal infections in children after the introduction of the 13-valent pneumococcal conjugate vaccine. *Pediatric Infectious Disease Journal, 32*(3), 203-207.

Kelly, T., Dillard, J. P., & Yother, J. (1994). Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infection and Immunity, 62*(5), 1813-1819.

Klevytska, A. M., Price, L. B., Schupp, J. M., Worsham, P. L., Wong, J., & Keim, P. (2001). Identification and characterization of Variable-Number Tandem Repeats in the *Yersinia pestis* genome. *Journal of Clinical Microbiology, 39*(9), 3179-3185. doi: 10.1128/JCM.39.9.3179-3185.2001.

Klugman, K. D. (1990). Pneumococcal resistance to antibiotics. *Clinical Microbiology Revised, 3*, 171-196.

Knutsen, E., Johnsborg, O., Quentin, Y., Claverys, J.-P., & Havarstein, L. S. (2006). BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. *Journal of Bacteriology, 188*(23), 8307-8312.

Ko, K. S., Baek, J. Y., & Song, J. H. (2013). Capsular gene sequences and genotypes of "serotype 6E" *Streptococcus pneumoniae* isolates. *Journal of Clinical Microbiology, 51*(10), 3395-3399.

Koeck, J., Nanjanpop-Lafourcade, B., Cade, S., Varon, E., Sangare, L., Valjevac, S., . . . Pourcel, C. (2005). Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. *BMC Microbiology, 5*(1), 66-74.

Koeuth, T., Versalovic, J., & Lupski, J. R. (1995). Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Research, 5*(4), 408-418.

Kolkman, M. A., Wakarchuk, W., Nuijten, P. J., & van der Zeijst, B. A. (1997). Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete cps locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. *Molecular Microbiology, 26*(1), 197-208.

Kroll, J. S., Loynds, B., Brophy, L. N., & Moxon, E. R. (1990). The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Molecular Microbiology, 4*(11), 1853-1862. doi: 10.1111/j.1365-2958.1990.tb02034.

Lawrence, E. R., Griffiths, D. B., Martin, S. A., George, R. C., & Hall, L. M. C. (2003). Evaluation of semiautomated multiplex PCR assay for determination of *Streptococcus pneumoniae* serotypes and serogroups. *Journal of Clinical Microbiology, 41*(2), 601-607.

Lehmann, D., Willis, J., Moore, H. C., Giele, C., Murphy, D., Keil, A. D., . . . Vaccine Impact Surveillance. (2010). The changing epidemiology of invasive pneumococcal disease in Aboriginal and non-Aboriginal Western Australians from 1997 through 2007 and emergence of nonvaccine serotypes. *Clinical Infectious Diseases, 50*(11), 1477-1486.

Lindstedt, B. (2005). Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis, 26*(13), 2567-2582.

Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., Le Doujet, C., . . . Vergnaud, G. (2006). Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci Multiple Locus Variable-Number Tandem Repeats Analysis. *BMC Microbiology, 6*(1), 33-33. doi: 10.1186/1471-2180-6-33.

Lott, T. J., Holloway, B. P., Logan, D. A., Fundyga, R., & Arnold, J. (1999). Towards understanding the evolution of the human commensal yeast *Candida albicans*. *Microbiology, 145*(5), 1137-1143. doi: 10.1099/13500872-145-5-1137.

Lu, Y.-J., Zhang, F., Sayeed, S., Thompson, C. M., Szu, S., Anderson, P. W., & Malley, R. (2012). A bivalent vaccine to protect against *Streptococcus pneumoniae* and *Salmonella typhi*. *Vaccine, 30*(23), 3405-3412.

Macleod, C. M., Hodges, R. G., Heidelberger, M., & Bernhard, W. G. (1945). Prevention of pneumococcal pneumonia by immunization with specific capsular polysaccharides. *The Journal of Experimental Medicine, 82*(6), 445-465.

Mak, D. B. (2004). Invasive pneumococcal disease in the Kimberley, 1995-2001. *The Australian Journal of Rural Health, 12*(6), 237-237.

Malachowa, N., Sabat, A., Gniadkowski, M., Krzyszton-Russjan, J., Empel, J., Miedzobrodzki, J., . . . Hryniewicz, W. (2005). Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, *spa* typing, and multilocus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. *Journal of Clinical Microbiology, 43*(7), 3095-3100.

Malley, R., & Anderson, P. W. (2012). Serotype-independent pneumococcal experimental vaccines that induce cellular as well as humoral immunity. *Proceedings of the National Academy of Sciences of the United States of America, 109*(10), 3623-3627.

Marchese, A., Esposito, S., Coppo, E., Rossi, G. A., Tozzi, A., Romano, M., . . . Principi, N. (2011). Detection of *Streptococcus pneumoniae* and identification of pneumococcal serotypes by Real-Time polymerase chain reaction using blood samples from Italian children ≤5 years of age with community-acquired pneumonia. *Microbial Drug Resistance, 17*(3), 419-424.

Marks, L. R., Reddinger, R. M., & Hakansson, A. P. (2012). High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *MBio, 3*(5). doi:10.1128/mBio.00200-12.

Marsh, R. L., Smith-Vaughan, H., Beissbarth, J., Hare, K., Kennedy, M., Wigger, C., . . . Leach, A. J. (2007). Molecular characterisation of pneumococcal serotype 16F: established predominant carriage and otitis media serotype in the 7vPCV era. *Vaccine, 25*(13), 2434-2436.

Martin, B., Humbert, O., Camara, M., Guenzi, E., Walker, J., Mitchell, T., . . . Hakenbeck, R. (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Research, 20*(13), 3479-3483.

Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kaltoft, M. S., Reeves, P. R., . . . Spratt, B. G. (2007). Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *Journal of Bacteriology, 189*(21), 7841-7855.

Mavroidi, A., Godoy, D., Aanensen, D. M., Robinson, D. A., Hollingshead, S. K., & Spratt, B. G. (2004). Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. *Journal of Bacteriology, 186*(24), 8181-8192.

Meats, E., Brueggemann, A. B., Enright, M. C., Sleeman, K., Griffiths, D. T., Crook, D. W., & Spratt, B. G. (2003). Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology, 41*(1), 386-392.

Mehr, S., & Wood, N. (2012). *Streptococcus pneumoniae*–a review of carriage, infection, serotype replacement and vaccination. *Paediatric Respiratory Reviews, 13*(4), 258-264.

McAuliffe, L., Ayling, R. D., & Nicholas, R. A. J. (2007). Identification and characterization of variable-number tandem-repeat markers for the molecular epidemiological analysis of *Mycoplasma mycoides* subspecies mycoides SC. *FEMS Microbiology Letters, 276*(2), 181.

McChlery, S. M., Scott, K. J., & Clarke, S. C. (2005). Clonal analysis of invasive pneumococcal isolates in Scotland and coverage of serotypes by the licensed conjugate polysaccharide

pneumococcal vaccine: possible implications for UK vaccine policy. *European Journal of Clinical Microbiology and Infectious Diseases, 24*(4), 262-267.

McGee, L., McDougal, L., Zhou, J., Spratt, B. G., Tenover, F. C., George, R., . . . Klugman, K. P. (2001). Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *Journal of Clinical Microbiology, 39*(7), 2565-2571.

McIntosh, E. D. G., & Reinert, R. R. (2011). Global prevailing and emerging pediatric pneumococcal serotypes. *Expert Review of Vaccines, 10*(1), 109-129.

Miernyk, K., Debyle, C., Harker-Jones, M., Hummel, K. B., Hennessy, T., Wenger, J., & Rudolph, K. (2011). Serotyping of *Streptococcus pneumoniae* isolates from nasopharyngeal samples: use of an algorithm combining microbiologic, serologic, and sequential multiplex PCR techniques. *Journal of Clinical Microbiology, 49*(9), 3209-3214.

Miller, E., Andrews, N. J., Waight, P. A., Slack, M. P. E., & George, R. C. (2011). Effectiveness of the new serotypes in the 13-valent pneumococcal conjugate vaccine. *Vaccine, 29*(49), 9127-9131.

Moberley, S., Holden, J., Tatham, D. P., & Andrews, R. M. (2013). Vaccines for preventing pneumococcal infection in adults. *Cochrane Database System Review, 1*.

Moffitt, K. L., Gierahn, T. M., Lu, Y.-J., Gouveia, P., Alderson, M., Flechtner, J. B., . . . Malley, R. (2011). TH17-based vaccine design for prevention of *Streptococcus pneumoniae* colonization. *Cell Host and Microbe, 9*(2), 158-165.

Morona, J. K., Morona, R., & Paton, J. C. (1999). Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. *Journal of Bacteriology, 181*(17), 5355-5364.

Morona, J. K., Miller, D. C., Coffey, T. J., Vindurampulle, C. J., Spratt, B. G., Morona, R., & Paton, J. C. (1999). Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 23F. *Microbiology, 145*(4), 781-789.

Moore, C. E., Peacock, S. J., Sengduangphachanh, A., Thaojaikong, T., Sirisouk, J., Foster, D., . . . Newton, P. N. (2010). Enhanced determination of *Streptococcus pneumoniae* serotypes associated with invasive disease in Laos by using a real-time polymerase chain reaction serotyping assay with cerebrospinal fluid. *The American Journal of Tropical Medicine and Hygiene, 83*(3), 451-457.

Moore, M. R., Gertz, J. R. E., Woodbury, R. L., Barkocy-Gallagher, G. A., Schaffner, W., Lexau, C., . . . Beall, B. (2008). Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. *The Journal of Infectious Diseases, 197*(7), 1016-1027.

Moore, M. R., Link-Gelles, R., Schaffner, W., Lynfield, R., Lexau, C., Bennett, N. M., . . . Whitney, C. G. (2015). Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. *The Lancet Infectious Diseases* (0). doi: http://dx.doi.org/10.1016/S1473-3099(14)71081-3.

Murray, J., Agocs, M., Serhan, F., Singh, S., Deloria-Knoll, M., O'Brien, K., . . . Cherian, T. (2014). Global invasive bacterial vaccine-preventable diseases surveillance - 2008-2014. *Morbidity and Mortality Weekly Report, 63*, 1159-1162.

NCIRS (Producer). (2015). Significant events in pneumococcal vaccination practice in Australia. Retrieved from: http://www.ncirs.edu.au/immunisation/history/Pneumococcal-history-March-2015.pdf.

Neufeld, F. (1902). Ueber die Agglutination der Pneumokokken und uber die Theorien der Agglutination. *Zeitschrift fur Hygiene Infektionskrankheiten*, 54-72.

Nesin, M., Ramirez, M., & Tomasz, A. (1998). Capsular transformation of a multidrug-resistant *Streptococcus pneumoniae in vivo*. *The Journal of Infectious Diseases, 177*(3), 707-713.

Njanpop Lafourcade, B.-M., Sanou, O., van der Linden, M., Levina, N., Karanfil, M., Yaro, S., . . . Mueller, J. E. (2010). Serotyping pneumococcal meningitis cases in the African meningitis belt by use of multiplex PCR with cerebrospinal fluid. *Journal of Clinical Microbiology, 48*(2), 612-614.

NNDSS. (2015). Disease notification rates, QLD, 1991 to 2012 and year-to-date notifications for 2015. Retrieved 12 January, 2015, from http://www9.health.gov.au/cda/source/cda-index.cfm.

Nunes, S., Sá-Leão, R., Carriço, J., Alves, C. R., Mato, R., Avô, A. B., . . . de Lencastre, H. (2005). Trends in drug resistance, serotypes, and molecular types of *Streptococcus pneumoniae* colonizing preschool-age children attending day care centers in Lisbon, Portugal: a summary of 4 years of annual surveillance. *Journal of Clinical Microbiology, 43*(3), 1285-1293.

O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., . . . Cherian, T. (2009). Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet, 374*(9693), 893-902.

Obszańska, K., Borek, A. L., Izdebski, R., Hryniewicz, W., & Sitkiewicz, I. (2011). Multilocus variable number tandem repeat analysis (MLVA) of *Streptococcus pyogenes*. *Journal of Microbiological Methods, 87*(2), 143-149.

Oftadeh, S., Gidding, H. F., & Gilbert, G. L. (2013). Laboratory surveillance of invasive pneumococcal disease in New South Wales, Australia, before and after introduction of 7-valent conjugate vaccine: reduced disease, but not antibiotic resistance rates. *Epidemiology and Infection, 141*(9), 1797-1806.

Oggioni, M. R., & Claverys, J. P. (1999). Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. *Microbiology, 145*(10), 2647.

Oliver, M. B., Van Der Linden, M. P. G., Küntzel, S. A., Saad, J. S., & Nahm, M. H. (2013). Discovery of *Streptococcus pneumoniae* serotype 6 variants with glycosyltransferases synthesizing two differing repeating units. *Journal of Biological Chemistry, 288*(36), 25976-25985.

O'Meara, T. R., & Alspaugh, J. A. (2012). The *Cryptococcus neoformans* capsule: a sword and a shield. *Clinical Microbiology Reviews, 25*(3), 387-408. doi: 10.1128/CMR.00001-12.

O'Riordan, K., & Lee, J. C. (2004). *Staphylococcus aureus* capsular polysaccharides. *Clinical Microbiology Reviews, 17*(1), 218-234. doi: 10.1128/CMR.17.1.218-234.2004.

Pai, R., Moore, M. R., Pilishvili, T., Gertz, R. E., Whitney, C. G., Beall, B., & Active Bacterial Core Surveillance, T. (2005). Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. *The Journal of Infectious Diseases, 192*(11), 1988-1995.

Pai, R., Gertz, R. E., & Beall, B. (2006). Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *Journal of Clinical Microbiology, 44*(1), 124-131.

Park, I. H., Geno, K. A., Yu, J., Oliver, M. B., Kim, K.-H., & Nahm, M. H. (2015). Genetic, biochemical, and serological characterization of a new pneumococcal serotype, 6H, and generation of a pneumococcal strain producing three different capsular repeat units. *Clinical and Vaccine Immunology , 22*(3), 313-318.

Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C. C., & Nahm, M. H. (2007). Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology, 45*(4), 1225-1233.

Picazo, J., Ruiz-Contreras, J., Casado-Flores, J., Giangaspro, E., Garcia-de-Miguel, M.-J., Hernandez-Sampelayo, T., . . . Mendez, C. (2013). Impact of introduction of conjugate vaccine in the vaccination schedule on the incidence of pediatric invasive pneumococcal disease requiring hospitalization in Madrid 2007 to 2011. *Pediatric Infectious Disease Journal, 32*(6), 656-661.

Pichon, B., Moyce, L., Sheppard, C., Slack, M., Turbitt, D., Pebody, R., . . . George, R. (2010). Molecular typing of pneumococci for investigation of linked cases of invasive pneumococcal disease. *Journal of Clinical Microbiology, 48*(5), 1926-1928.

Pimenta, F. C., Beall, B., Roundtree, A., Soysal, A., Bakir, M., du Plessis, M., . . . Carvalho, M. d. G. (2013). Sequential triplex real-time PCR assay for detecting 21 pneumococcal capsular serotypes that account for a high global disease burden. *Journal of Clinical Microbiology, 51*(2), 647-652.

Platt, S. (2015). VNTR DIversity and Confidence Extractor. Retrieved 13 August, 2013, from http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl.

PMEN. (2015). Pneumococcal Molecular Epidemiology Network (PMEN). Retrieved 30 June, 2015, from http://web1.sph.emory.edu/PMEN/pmen_criteria.html..html.

Poehling, K. A., Talbot, T. R., Griffin, M. R., Craig, A. S., Whitney, C. G., Zell, E., . . . Schaffner, W. (2006). Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. *Journal of the American Medical Association, 295*(14), 1668-1674.

Pourcel, C., Minandri, F., Hauck, Y., D'Arezzo, S., Imperi, F., Vergnaud, G., & Visca, P. (2011). Identification of Variable-Number Tandem-Repeat (VNTR) Sequences in *Acinetobacter baumannii* and interlaboratory validation of an optimized Multiple-Locus VNTR Analysis Typing scheme. *Journal of Clinical Microbiology, 49*(2), 539-548.

Price, E. P., Thiruvenkataswamy, V., Mickan, L., Unicomb, L., Rios, R. E., Huygens, F., & Giffard, P. M. (2006). Genotyping of *Campylobacter jejuni* using seven single-nucleotide polymorphisms in combination with *flaA* short variable region sequencing. *Journal of Medical Microbiology, 55*(8), 1061-1070.

Prudhomme, M., Martin, B., Mejean, V., & Claverys, J. P. (1989). Nucleotide sequence of the *Streptococcus pneumoniae hexB* mismatch repair gene: homology of *HexB* to *MutL* of *Salmonella typhimurium* and to *PMS1* of *Saccharomyces cerevisiae*. *Journal of Bacteriology, 171*(10), 5332-5338.

Qu, W., Shen, Z., & Zhang, C. (2009). PriDimerCheck. Retrieved 25 June, 2012, from http://biocompute.bmi.ac.cn/MPprimer/primer_dimer.html.

Queensland Government. (2015). Queensland Population Counter. Retrieved 17 February, 2015, from http://www.oesr.qld.gov.au/products/briefs/pop-growth-qld/qld-pop-counter.php.

Rakov, A. V., Ubukata, K., & Ashley Robinson, D. (2011). Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus *boxB* sequence typing. *Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases, 11*(8), 1929-1939.

Randle, E., Ninis, N., & Inwald, D. (2011). Invasive pneumococcal disease. *Archives of Disease in Childhood - Education and Practice, 96*, 183-190.

Rathnayake, I. U., Hargreaves, M., & Huygens, F. (2011). Genotyping of *Enterococcus faecalis* and *Enterococcus faecium* isolates by use of a set of eight single nucleotide polymorphisms. *Journal of Clinical Microbiology, 49*(1), 367-372.

Rayner, R. E., Savill, J., Hafner, L. M., & Huygens, F. (2015). Genotyping *Streptococcus pneumoniae*. *Future Microbiology*, 10(4): 653-664.

Rayner, R. E., Savill, J., Hafner, L. M., & Huygens, F. (2015). Modified MLVA for genotyping Queensland invasive *Streptococcus pneumoniae*. *Plos One, 10*(4), e0121870. doi:10.1371/journal.pone.0121870.

Reingold, A., Hadler, J., Farley, M. M., & Harrison, L. (2005). Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease – United States, 1998-2003. *Journal of the American Medical Association, 294*(16), 2022-2026.

Riley, I. D., Andrews, M., Howard, R., Tarr, P. I., Pfeiffer, M., & Challands, P. (1977). Immunization with polyvalent pneumococcal vaccine. Reduction of adult respiratory mortality in a New Guinea Highlands community. *Lancet, 1*(8026), 1338-1341.

Robertson, G. A., Thiruvenkataswamy, V., Shilling, H., Price, E. P., Huygens, F., Henskens, F. A., & Giffard, P. M. (2004). Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. *Journal of Medical Microbiology, 53*(Pt 1), 35-45.

Robbins, J., & Schneerson, R. (1983). Planning for a 2nd (23-valent) generation pneumococcal vaccine – with special reference to new developments in our understanding of the

structure and biology of polysaccharides. *Clinical Respiratory Physiology, 19*(2), 215-226.

Roche, P. W., Krause, V., Cook, H., Barralet, J., Coleman, D., Sweeny, A., . . . Murphy, D. (2008). Invasive pneumococcal disease in Australia, 2006. *Communicable Diseases Intelligence, 32*(1), 18-30.

Roche, P., Krause, V., Cook, H., Bartlett, M., Coleman, D., Davis, C., . . . Enhanced Invasive Pneumococcal Diseases Surveillance Working. (2007). Invasive pneumococcal disease in Australia, 2005. *Communicable Diseases Intelligence, 31*(1), 86-100.

Roche, P., Krause, V., Andrews, R., Carter, L., Coleman, D., Cook, H., . . . Pneumococcal Working party of the Communicable Diseases Network. (2003). Invasive pneumococcal disease in Australia, 2002. *Communicable Diseases Intelligence, 27*(4), 466-477.

Rodgers, G. L., & Klugman, K. P. (2011). The future of pneumococcal disease prevention. *Vaccine, 29*(3), 43-48.

Sabharwal, V., Stevenson, A., Figueira, M., Orthopoulos, G., Trzciński, K., & Pelton, S. I. (2014). Capsular switching as a strategy to increase pneumococcal virulence in experimental otitis media model. *Microbes and Infection, 16*(4), 292-299. doi: 10.1016/j.micinf.2013.12.002.

Saha, S. K., Luby, S. P., Santosham, M., Crook, D., Darmstadt, G. L., Baqui, A. H., . . . Arifeen, S. E. (2008). Identification of serotype in culture negative pneumococcal meningitis using sequential multiplex PCR: implication for surveillance and vaccine design. *PLoS One, 3*(10), e3576. doi:10.1371/journal.pone.0003576.

Saha, S., Modak, J. K., Naziat, H., Al-Emran, H., Chowdury, M., Islam, M., . . . Saha, S. K. (2015). Detection of co-colonization with *Streptococcus pneumoniae* by algorithmic use of conventional and molecular methods. *Vaccine, 33*(5), 713-718. doi: 10.1016/j.vaccine.2014.11.040.

Saha, S. K., Rikitomi, N., Biswas, D., Watanabe, K., Ruhulamin, M., Ahmed, K., . . . Nagatake, T. (1997). Serotypes of *Streptococcus pneumoniae* causing invasive childhood infections in Bangladesh, 1992 to 1995. *Journal of Clinical Microbiology, 35*(3), 785-787.

Sá-Leão, R., Pinto, F., Aguiar, S., Nunes, S., Carriço, J. A., Frazão, N., . . . Ramirez, M. (2011). Analysis of invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines reveals heterogeneous behavior of clones expressing the same serotype. *Journal of Clinical Microbiology, 49*(4), 1369-1375.

Salter, S. J., Hinds, J., Gould, K. A., Lambertsen, L., Hanage, W., Antonio, M., . . . Bentley, S. D. (2012). Variation at the capsule locus, cps, of mistyped and non-typable *Streptococcus pneumoniae* isolates. *Microbiology (United Kingdom), 158*(6), 1560-1569.

Sanchez-Beato, A. R., Lopez, R., & Garcia, J. L. (1998). Molecular characterization of PcpA: a novel choline-binding protein of *Streptococcus pneumoniae*. *FEMS Microbiology Letters, 164*, 207-214.

Sandgren, A., Sjostrom, K., Olsson-Liljequist, B., Christensson, B., Samuelsson, A., Kronvall, G., & Henriques Normark, B. (2004). Effect of clonal and serotype-specific properties on the

invasive capacity of *Streptococcus pneumoniae*. *The Journal of Infectious Diseases, 189*(5), 785-796.

Sarkar, S., Ulett, G. C., Totsika, M., Phan, M.-D., & Schembri, M. A. (2014). Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. *PloS One, 9*(4), e94786. doi: 10.1371/journal.pone.0094786.

Scott, J. A. G., Hall, A. J., Dagan, R., Dixon, J. M. S., Eykyn, S. J., Fenoll, A., . . . Smart, L. E. (1996). Serogroup-specific epidemiology of *Streptococcus pneumoniae*: associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clinical Infectious Diseases, 22*(6), 973-981.

Schouls, L., & Van der Heide, H. (2012). *Streptococcus pneumoniae*. Retrieved April, 12, 2012, from http://www.mlva.net/spneumoniae/default.asp.

Selvey, C. (2011). Information on the Prevenar 13 Supplementary Dose Program. Retrieved May, 29, 2012, from http://www.seagp.org.au/documents/Services/PracticeSupport/PublicHealth/20110602_LETTER_QHIP_to_providers_JS.pdf.

Severiano, A., Pinto, F. R., Ramirez, M., & Carriço, J. A. (2011). Adjusted Wallace coefficient as a measure of congruence between typing methods. *Journal of Clinical Microbiology, 49*(11), 3997-4000.

Sharma, S. K., Roumanes, D., Almudevar, A., Mosmann, T. R., & Pichichero, M. E. (2013). CD4+ T-cell responses among adults and young children in response to *Streptococcus pneumoniae* and *Haemophilus influenzae* vaccine candidate protein antigens. *Vaccine, 31*(30), 3090-3097.

Sheludchenko, M. S., Huygens, F., & Hargreaves, M. H. (2010). Highly discriminatory single-nucleotide polymorphism interrogation of *Escherichia coli* by use of allele-specific real-time PCR and eBURST analysis. *Applied and Environmental Microbiology, 76*(13), 4337-4345.

Siira, L., Kaijalainen, T., Lambertsen, L., Nahm, M. H., Toropainen, M., & Virolainen, A. (2012). From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. *Journal of Clinical Microbiology, 50*(8), 2727-2731.

Simões, A. S., Pereira, L., Nunes, S., Brito-Avô, A., Lencastre, H. d., & Sá-Leão, R. (2011). Clonal evolution leading to maintenance of antibiotic resistance rates among colonizing pneumococci in the PCV7 era in Portugal. *Journal of Clinical Microbiology, 49*(8), 2810-2817.

Singleton, R., Wenger, J. D., Klejka, J. A., Bulkow, L. R., Thompson, A., Sarkozy, D., . . . Scott, D. A. (2013). The 13-valent pneumococcal conjugate vaccine for invasive pneumococcal disease in Alaska native children: results of a clinical trial. *Paediatric Respiratory Reviews, 32*(3), 257-263.

Singleton, R. J., Hennessy, T. W., Bulkow, L. R., Hammitt, L. L., Zulz, T., Hurlburt, D. A., . . . Parkinson, A. (2007). Invasive pneumococcal disease caused by nonvaccine serotypes

among Alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage. *JAMA, 297*(16), 1784-1792.

Skinner, J. M., Indrawati, L., Cannon, J., Blue, J., Winters, M., Macnair, J., . . . Heinrichs, J. H. (2011). Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine, 29*(48), 8870 – 8876.

Slaon-Gardner, T., Stirzaker, S., Knuckey, D., Pennington, K., Knope, K., Fitzsimmons, G., . . . Group, N. A. R. W. (2011). Australia's notifiable disease status, 2009: annual report of the National Notifiable Diseases Surveillance System. *Communicable Diseases Intelligence, 35*(2), 61-131.

Spurgiesz, R. S., Quitugua, T. N., Smith, K. L., Schupp, J., Palmer, E. G., Cox, R. A., & Keim, P. (2003). Molecular typing of *Mycobacterium tuberculosis* by using nine novel Variable-Number Tandem Repeats across the Beijing family and low-copy-number IS6110 isolates. *Journal of Clinical Microbiology, 41*(9), 4224-4230. doi: 10.1128/JCM.41.9.4224-4230.2003.

Staples, M., Graham, R., Ariotti, L., Cook, H., Krause, V., Giele, C., . . . Smith, H. (2012). *Molecular characterisation of an Australian serotype 1 Streptococcus pneumoniae outbreak*. Paper presented at the Australian Society of Microbiology, Brisbane, Queensland. Webpage retrieved from http://www.asm2012.org/abstract/70.asp.

Staples, M., Graham, R. M. A., Jennison, A. V., Ariotti, L., Hicks, V., Cook, H., . . . Smith, H. V. (2015). Molecular characterization of an Australian serotype 1 *Streptococcus pneumoniae* outbreak. *Epidemiology and Infection, 143*(2), 325-333.

Statens Serum Institut. (2013). Pneumococcus antisera. Retrieved 20 March, 2014, from http://www.ssi.dk/English/SSI%20Diagnostica/Products%20from%20SSI%20Diagnostica/Antisera_antibodies/Pneumococcus%20antisera.aspx.

Stephens, A. J., Huygens, F., & Giffard, P. M. (2007). Systematic derivation of marker sets for *Staphylococcal* cassette chromosome *mec* typing. *Antimicrobial Agents and Chemotherapy, 51*(8), 2954-2964.

Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). Two-component signal transduction. *Annual Review of Biochemistry, 69*(1), 183-215.

Straume, D., Stamsås, G. A., & Håvarstein, L. S. (2014). Natural transformation and genome evolution in *Streptococcus pneumoniae*. *Infection, Genetics and Evolution, In Press*. doi:10.1016/j.meegid.2014.10.020.

Swartley, J. S., Marfin, A. A., Edupuganti, S., Cieslak, P., Perkins, B., Wenger, J. D., & Stephens, D. S. (1997). Capsule switching of *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America, 94*(1), 271-276. doi: 10.1073/pnas.94.1.271.

Swiatlo, E., & Ware, D. (2003). Novel vaccine strategies with protein antigens of *Streptococcus pneumoniae*. *FEMS immunology and medical microbiology, 38*(1), 1-7.

Tarrago, D., Fenoll, A., Sanchez-Tatay, D., Arroyo, L. A., Munoz-Almagro, C., Esteva, C., . . . Obando, I. (2008). Identification of pneumococcal serotypes from culture-negative clinical specimens by novel real-time PCR. *Clinical Microbiology and Infection, 14*(9), 828-828.

Temime, L., Boelle, P.-Y., Opatowski, L., & Guillemot, D. (2008). Impact of capsular switch on invasive pneumococcal disease incidence in a vaccinated population. *PLoS One, 3*(9), e3244. doi:10.1371/journal.pone.0003244.

Todar, K. (2012). *Streptococcus pneumoniae*. Retrieved April, 24, 2012, from http://textbookofbacteriology.net/S.pneumoniae.html.

Tomita, Y., Okamoto, A., Yamada, K., Yagi, T., Hasegawa, Y., & Ohta, M. (2011). A new microarray system to detect *Streptococcus pneumoniae* serotypes. *Journal of Biomedicine and Biotechnology, 2011*, 352736-352721.

Torzillo, P. J., Morey, F., Gratten, M., Murphy, D., Matters, R., & Dixon, J. (2007). Changing epidemiology of invasive pneumococcal disease in central Australia prior to conjugate vaccine: a 16-year study. *Vaccine, 25*(13), 2375-2378.

Tran, T., Clarke, M., & Marshall, H. (2012). Invasive pneumococcal disease: impact of pneumococcal vaccination in South Australian children. *International Journal of Infectious Diseases, 16*, E309-E309. doi:10.1016/j.ijid.2012.05.1002.

Trappetti, C., Erika van der, M., Amin, Z., Potter, A. J., Chen, A. Y., Paula, M. v. M., . . . Paton, J. C. (2013). Site of Isolation Determines Biofilm Formation and Virulence Phenotypes of *Streptococcus pneumoniae* Serotype 3 Clinical Isolates. *Infection and immunity, 81*(2), 505-513.

Trück, J., Snape, M. D., Tatangeli, F., Voysey, M., Yu, L.-M., Faust, S. N., . . . Pollard, A. J. (2014). Pneumococcal serotype-specific antibodies persist through early childhood after infant immunization: follow-up from a randomized controlled trial. *PLoS One, 9*(3). doi:10.1371/journal.pone.0091413.

Tyrrell, G. J., Lovgren, M., Chui, N., Minion, J., Garg, S., Kellner, J. D., & Marrie, T. J. (2009). Serotypes and antimicrobial susceptibilities of invasive *Streptococcus pneumoniae* pre- and post-seven valent pneumococcal conjugate vaccine introduction in Alberta, Canada, 2000–2006. *Vaccine, 27*(27), 3553-3560.

Van Belkum, A., Sluijuter, M., de Groot, R., Verbrugh, H., & Hermans, P. W. (1996). Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. *Journal of Clinical Microbiology, 34*(5), 1176-1179.

Van Cuyck, H., Pichon, B., Leroy, P., Granger-Farbos, A., Underwood, A., Soullie, B., & Koeck, J. L. (2012). Multiple-Locus Variable-Number Tandem-Repeat Analysis of *Streptococcus pneumoniae* and comparison with Multiple Loci Sequence Typing. *BMC Microbiology, 12*, 241-258.

Van der Linden, M., Reinert, R. R., Kern, W. V., & Imöhl, M. (2013). Epidemiology of serotype 19A isolates from invasive pneumococcal disease in German children. *BMC Infectious Diseases, 13*(1), 70.

Venkateswaran, P. S., Stanton, N., & Austrian, R. (1983). Type variation of strains of *Streptococcus pneumoniae* in capsular serogroup 15. *The Journal of Infectious Diseases, 147*(6), 1041-1054.

Verhoeven, D., Xu, Q., & Pichichero, M. E. (2014). Vaccination with a *Streptococcus pneumoniae* trivalent recombinant PcpA, PhtD and PlyD1 protein vaccine candidate protects against lethal pneumonia in an infant murine model. *Vaccine, 32*(26), 3205-3210.

Wang, X., Huang, B., Blair, B., Eglezos, S., & Bates, J. (2012). Selection of optimal combinations of loci by the Optimal Combination Finder computer program from a group of variable number tandem repeat loci for use in *Staphylococcus aureus* food poisoning case investigations. *Journal of Medical Microbiology, 61*(5), 631.

Walker, J. A., Allen, R. L., Falmagne, P., Johnson, M. K., & Boulnois, G. J. (1987). Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of *Streptococcus pneumoniae*. *Infection and Immunity, 55*(5), 1184-1189.

Wallace, D. L. (1983). A Method for Comparing Two Hierarchical Complexings: Comment. *Journal of the American Statistical Association, 78*(383), 569-576.

Weatherholtz, R., Millar, E. V., Moulton, L. H., Reid, R., Rudolph, K., Santosham, M., & O'Brien, K. L. (2010). Invasive pneumococcal disease a decade after pneumococcal conjugate vaccine use in an American Indian population at high risk for disease. *Clinical Infectious Diseases, 50*(9), 1238-1246.

Weinberger, D. M., Malley, R., & Lipsitch, M. (2011). Serotype replacement in disease after pneumococcal vaccination. *Lancet, 378*(9807), 1962-1973.

Whitney, C. G., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Lynfield, R., . . . Schuchat, A. (2003). Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *The New England Journal of Medicine, 348*(18), 1737-1746.

Wickstead, B., Ersfeld, M., & Gull, K. (2003). Repetitive elements in genomes of parasitic protozoa. *Microbiology and Molecular Biology Reviews, 67*(3), 360-360.

Williams, S. R., Mernagh, P. J., Lee, M. H. T., & Tan, J. T. (2011). Changing epidemiology of invasive pneumococcal disease in Australian children after introduction of a 7-valent pneumococcal conjugate vaccine. *Medical Journal of Australia, 194*(3), 116-120.

Wright, A. E., Morgan, W. P., Colebrook, L., & Dodgson, R. W. (1914). Prophylactic inoculation against pneumococcus infections and on the results which have been achieved by it. *Lancet, 1*, 87-95.

Wu, K., Yao, R., Wang, H., Pang, D., Liu, Y., Xu, H., . . . Yin, Y. (2014). Mucosal and systemic immunizations with a novel attenuated pneumococcal vaccine candidate confer serotype independent protection against *Streptococcus pneumoniae* in mice. *Vaccine, 32*(33), 4179-4188.

Wyres, K. L., Klugman, K. P., Parkhill, J., Hakenbeck, R., Bentley, S. D., Brueggemann, A. B., . . . Beall, B. W. (2013). Pneumococcal capsular switching: a historical perspective. *The Journal of Infectious Diseases, 207*(3), 439-449.

Xue, S.-A., Jones, M. D., Lu, Q.-L., Middeldorp, J. M., & Griffin, B. E. (2003). Genetic diversity: frameshift mechanisms alter coding of a gene (Epstein-Barr Virus *LF3* gene) that contains multiple 102-base-pair direct sequence repeats. *Molecular and Cellular Biology, 23*(6), 2192-2201.

Yaro, S., Lourd, M., Traore, Y., Njanpop-Lafourcade, B. M., Sawadogo, A., Sangare, L., . . . Gessner, B. D. (2006). Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. *Clinical Infectious Diseases, 43*(6), 693-700.

Yu, J., Lin, J., Kim, K.-H., Benjamin, J. W. H., & Nahm, M. H. (2011). Development of an automated and multiplexed serotyping assay for *Streptococcus pneumoniae*. *Clinical and Vaccine Immunology, 18*(11), 1900-1907.

Zhou, F., Kong, F., Tong, Z., & Gilbert, G. L. (2007). Identification of less-common *Streptococcus pneumoniae* serotypes by a multiplex PCR-based reverse line blot hybridization assay. *Journal of Clinical Microbiology, 45*(10), 3411-3415.

Ziebuhr, W., Krimmer, V., Rachid, S., Lößner, I., Götz, F., & Hacker, J. (1999). A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Molecular Microbiology, 32*(2), 345-356.

APPENDIX A

Appendix A1: Position of MLST housekeeping primers in five *S. pneumoniae* genomes.

	Position in *S. pneumoniae* genome				
Primer	G54	CGSP14	TIGR4	Hungary6-19A	R6
aroE F (17bp)	1283903- 1283887 (34.2, 0.002)	1352222- 1352206 (34.2, 0.002)	1299642- 1299626 (34.2, 0.002)	1395063- 1395047 (34.2, 0.002)	1232203- 1232187 (34.2, 0.002)
aroE R (22bp)	1283425- 1283441 (26.3, 1.2)	1351744- 1351761 (28.2, 0.3)	1299164- 1299181 (28.2, 0.3)	1394585- 1394602 (28.2, 0.3)	1231725- 1231742 (28.2, 0.3)
LENGTH	478	478	478	478	478
Gdh F (20)	1111556- 1111575 (32.2, 0.015)	1032033- 1032014 (32.2, 0.015)	1175299- 1175318 (32.2, 0.015)	1258597- 1258616 (32.2, 0.015)	1123456- 1123475 (32.2, 0.015)
Gdh R (18bp)	1112214- 1112198 (34.2, 0.003)	1031375- 1031391 (34.2, 0.003)	1175957- 1175941 (34.2, 0.003)	1259255- 1259239 (34.2, 0.003)	1124114- 1124098 (34.2, 0.003)
LENGTH	658	658	658	658	658
Gki F (20bp)	605585- 605604 (40.1, 6e-5)	633259- 633278 (40.1, 6e-5)	644274- 644293 (40.1, 6e-5)	719036- 719052 (26.3, 0.93)	600349- 600368 (40.1, 6e-5)
Gki R (17bp)	606210- 606194 (34.2, 0.002)	633884- 633868 (34.2, 0.002)	644899- 644883 (34.2, 0.002)	719658- 719642 (34.2, 0.002)	600974- 600958 (34.2, 0.002)
LENGTH	625	625	625	622	625
recP F (20bp)	1846894- 1846875 (40.1, 6e-5)	1983668- 1983649 (40.1, 6e-5)	1936947- 1936928 (40.1, 6e-5)	2009720- 2009701 (40.1, 6e-5)	1817870- 1817851 (40.1, 6e-5)
recP R (20bp)	1846326- 1846344 (22.3, 15)	1983100- 1983118 (22.3, 15)	1936379- 1936397 (22.3, 15)	2009152- 2009170 (22.3, 15)	1817302- 1817320 (22.3, 15)
LENGTH	568	568	568	568	568
Spi F (20bp)	362843- 362827 (26.3, 0.93)	398225- 398206 (40.1, 6e-5)	383032- 383016 (26.3, 0.93)	465369- 465350 (40.1, 6e-5)	364453- 364434 (40.1, 6e-5)
Spi R (20bp)	362287- 362306 (32.2, 0.015)	397666- 397685 (40.1, 6e-5)	382476- 382495 (32.2, 0.015)	464810- 464829 (40.1, 6e-5)	363894- 363913 (40.1, 6e-5)
LENGTH	556	559	556	559	559
Xpt F (20bp)	1668944- 1668963 (40.1, 6e-5)	1806225- 1806244 (40.1, 6e-5)	1755283- 1755302 (40.1, 6e-5)	1823498- 1823517 (32.2, 0.015)	1635326- 1635345 (40.1, 6e-5)
Xpt R (20bp)	1669515- 1669496 (40.1, 6e-5)	1806796- 1806777 (40.1, 6e-5)	1755854- 1755835 (40.1, 6e-5)	1824069- 1824050 (40.1, 6e-5)	1635897- 1635878 (40.1, 6e-5)
LENGTH	571	571	571	571	571
Ddl F (20bp)	1522003- 1521984 (40.1, 6e-5)	1615678- 1615659 (40.1, 6e-5)	1571797- 1571778 (40.1, 6e-5)	1642333- 1642314 (40.1, 6e-5)	1493017- 1492998 (40.1, 6e-5)
Ddl R (20bp)	1521491- 1521510 (24.3, 3.7)	1615166- 1615185 (32.2, 0.015)	1571285- 1571304 (32.2, 0.015)	1641821- 1641840 (32.2, 0.015)	1492505- 1492524 (32.2, 0.015)
LENGTH	512	512	512	512	512

Appendix A2: BOX primer position in five *S. pneumoniae* genomes.

The presence of a red-coloured base-pair in a primer indicates that it was absent when blasted against genomes. A substituted base-pair has the new base-pair in the brackets next to red base. Symbols for each genome are as follows: *R6, **Hung19A, ***G54, ^CGSP14, ^^TIGR4.

		Position in genome				
Primer	**Sequence**	**R6 (serotype 2)***	**Hungary19A-6 ** (19A)**	**G54 (19F)*****	**CGSP14 (14)^**	**TIGR4 (4)^^**
BOX-01f (ms40) 21-mer	CCAGAGACATTGATGAAGAGA	1611150-1611170 (4e-06, 42.1)	1802475-1802495 (4e-06, 42.1)	1647836-1647856 (4e-06, 42.1)	1785224-1785244 (4e-06, 42.1)	1731039-1731059 (42.1)
BOX-01r 20-mer	CGCTTTGATGAACTTGAGTT	1611476-1611457 (1 e -05, 40.1)	1802846-1802827 (1e-05, 40.1)	1648297-1648278 (1e-05, 40.1)	1785595-1785576 (1e-05, 40.1)	1731455-1731436 (40.1)
BOX-01	**LENGTH (bp)**	**326**	**371**	**461**	**371**	**377**
BOX-02f (ms32) 21-mer (** is reverse)	TTGCTTGGTACAGAAAACAAA	571757- 571777 (4e-06, 42.1)	689614-689634 (1e-05, 42.1)	576435-576455(1e-05, 42.1)	603709-603729 (1e-05, 42.1)	614986-615006 (42.1)
BOX-02r 21-mer (** is forward)	CCCCATAAAACCCTCCTTATA	572128-572108 (4 e -06, 42.1)	690030-690010 (4e-06, 42.1)	576851-576831 (4e-06, 42.1)	604125-604105 (4e-06, 42.1)	615447-615427 (42.1)
BOX-02	**LENGTH**	**371**	**416**	**416**	**416**	**421**
BOX-03f (ms15) 19-mer (** is reverse)	TCCAACACGACCTTTATCC	1579329-1579347 (1e-04, 38.2)	1763019-1763037 (1e-04, 38.2)	1608845-1608863 (5e-05, 38.2)	1741339-1741357 (1e-04, 38.2)	1695625-1695643 (38.2)
BOX-03r 20-mer (** is forward)	TTCAGTAAACCCAGCTCGTA	1579996-1579977 (1 e -05, 40.1)	1763686-1763667 (1e-05, 40.1)	1609376-1609357 (1e-05, 40.1)	1741781-1741762 (1e-05, 40.1)	1696336-1696317 (40.1)
BOX-03	**LENGTH**	**667**	**667**	**531**	**442**	**674**
BOX-04f (ms33) 21-mer (*** is reverse)	TGGGTAAAAGTAGACAGGACT	698458- 698478 (4 e -06, 42.1)	822567-822587 (4e-06, 42.1)	698658-698678 (4e-06, 42.1) not anneal?	730880-730900 (4e-06, 42.1)	
BOX-04r 25-mer (*** is forward)	CACTTCTACACTAGTTTGTAAAGCA	698736- 698712 (7e-08, 50.1)	824313-824289 (7e-08, 50.1)	699025-699002 (9e-08, 48.1)	731203-731179 (7e-08, 50.1)	
BOX-04	**LENGTH**	**278**	**1746**	**367**	**323**	
BOX 04 F2 21-mer	AGGGGATTTACCCACTACAAA	41592- 41612 (4 e -06, 42.1) repeated	823993 - 824013 (1e-05, 42.1) repeated	760958-760978 (4e-06, 42.1) repeated	793862-793882 (4e-06, 42.1) repeated	
BOX 4 Pr 25-mer	CACTTCTACACTAGTTTGTAAAGCA	698736- 698712 (7e-08, 50.1)	824313-824289 (7e-08, 50.1)	699025-699002 (9e-08, 48.1)	731203-731179 (7e-08, 50.1)	
BOX-04 (2)	**LENGTH**	**-**	**317**	**-**	**-**	
BOX 06 Nf**,***,^ 21-mer	GAAAAAGGTCAGGAGTAG(A)GTT	1911454-1911474 (8e-06, 42.1)	2117294-2117314 (7e-06, 42.1)	1946160-1946175 (0.007, 32.2) not anneal?	2077187-2077203 (0.002, 34.2) not anneal?	
BOX 06 R 19-mer	TCACTTGAGACAATCAGCC	1911736-1911718 (5 e -05, 38.2)	2117588-2117570(5e-05, 38.2)	1946453-1946435(5e-05, 38.2)	2077526-2077508 (5e-05, 38.2)	
BOX-06	**LENGTH**	**282**	**294**	**293**	**339**	

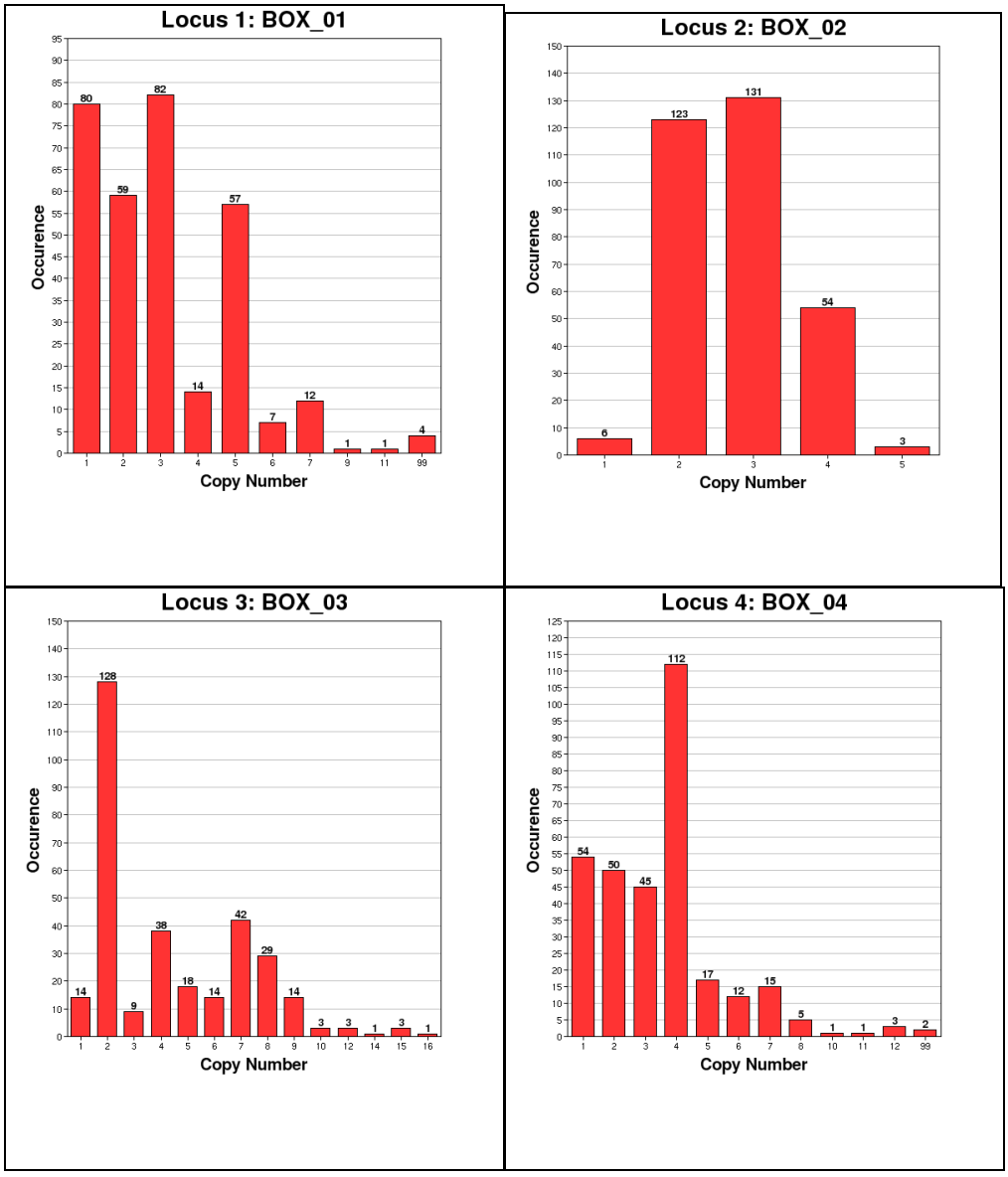
		Position in genome				
Primer	**Sequence**	**R6 (serotype 2)***	**Hungary19A-6 ** (19A)**	**G54 (19F)*****	**CGSP14 (14)^**	**TIGR4 (4)^^**
BOX 06 F2 22-mer	TTATGATTTTTGTCATTTTGCA	1911429-1911450 (1 e -06, 44.1)	2117269-2117290 (1e-06, 44.1)	1946134-1946155 (1e-06, 44.1)	2077162-2077183 (1e-06, 44.1)	
(367)B(303)OX 6 Nr2 24-mer	GAAATCTTTGAAAAACTAGGATTT	1911706-1911683 (3e-07, 48.1)	2117558-2117535 (3e-07, 48.1)	1946423-1946400 (3e-07, 48.1)	2077496-2077473 (3e-07, 48.1)	
BOX-06 (2)	**LENGTH**	**277**	**289**	**289**	**334**	
B10f 20-mer	TATGACGCGTATGGAAGTGC	587873 - 587892 (40.1)	705790 - 705809 (40.1)	592613 - 592632 (40.1)	619880 - 619899 (40.1)	631207 - 631226 (40.1)
B10r* 23-mer	GATTGTGTAGGTCATGC(A)TCTGCT	588335 - 588313 (46.1)	706432 - 706410 (38.2)	593075 - 593053 (38.2)	620432 - 620410 (38.2)	631669 - 631647 (46.1)
BOX -10	**LENGTH**	**462**	**642**	**462**	**552**	**462**
BOX 11 Vf 23-mer* **	TCCAATAATGACAGG(A)TTTTCCTC	411593- 411615 (9e-07, 46.1)	512935-512957 (2e-04, 38.2)	411736-411758 (3e-07, 46.1)	447095-447117 (9e-07, 46.1)	
BOX 11 R 21-mer	TTCCAATCTACGCCTTTGAAG	412189- 412169 (4 e -06, 42.1)	513529-513509 (4e-06, 42.1)	412285-412265 (4e-06, 42.1)	447644-447624 (4e-06, 42.1)	
BOX-11	**LENGTH**	**596**	**594**	**549**	**549**	
BOX 12 Pf 19-mer (** is reverse)	TTGCCCTTTTCATCTTCGA	1350097-1350115 (2e-04, 38.2)	1501035-1501053 (2e-04, 38.2)	1389836-1389854 (2e-04, 38.2)	1472944-1472962 (2e-04, 38.2)	
BOX 12 R 18-mer (** is forward)	CAGCAACCATTGAAACGC	1350841-1350824 (2 e -04, 36.2)	1501824-1501807 (2e-04, 36.2)	1390400-1390383 (2e-04, 36.2)	1473823-1473806 (2e-04, 36.2)	
BOX-12	**LENGTH**	**744**	**789**	**564**	**879**	
BOX-13f (ms25) 20-mer	TCGCCTTTGCTAATCAAACA	101033-101052 (5e-05, 40.1)	178817-178836 (5e-05, 40.1)	103496-103515 (5e-05, 40.1)	101226-101245 (5e-05, 40.1)	
BOX-13r 23-mer	CTGATTATATCGCTCACAAACCC	101512-101490 (3e-07, 46.1)	179161-179139(3e-07, 46.1)	103974-103952 (3e-07, 46.1)	101702-101680 (3e-07, 46.1)	
BOX-13	**LENGTH**	**479**	**344**	**478**	**476**	
Spneu31f 21-mer	CTGGAATAGTCCATCGAGCTT	556641- 556661	673012-673032	561723-561743	588988-589008	598431 - 598451 (42.1)
Spneu31r 21-mer	CGAAGGTAAATGTGAACAAAC	557234- 557214	673290-673270	561956-561936	589221-589201	598709 - 598689 (42.1)
Spneu31	**LENGTH**	**593**	**278**	**233**	**233**	**278**
Sp BOX-12Pf NEW 19-mer (** is reverse)	GATTGCCCTTTTCATCTTC	1350095 -1350113 (38.2)	1501033 -1501051 (38.2)	1389834 - 1389852 (38.2)	1472942 -1472960 (38.2)	1424174 -1424192 (38.2)
Sp BOX-12r NEW 18-mer (** is forward)	CACAGCAACCATTGAAAC	1350843 - 1350826 (36.2)	1501826 - 1501809 (36.2)	1390402 - 1390385 (36.2)	1473825 - 1473808 (36.2)	1424876 - 1424860 (34.2)
QHFSS BOX12		**748**	**793**	**568**	**883**	**702**
Spneu15f 20-mer	TCCAACACGACCTTTATCCA	1579329-1579348	1763019-1763038	1608845-1608864	1741339-1741358	1695625 - 1695644 (40.1)
Spneu15r 20-mer	AGCCTTGCTCCTCATCTTGA	1579835-1579816	1763525-1763506	1609215-1609196	1741620-1741601	1696175 - 1696156 (40.1)
Spneu15		**468**	**468**	**332**	**243**	**531**

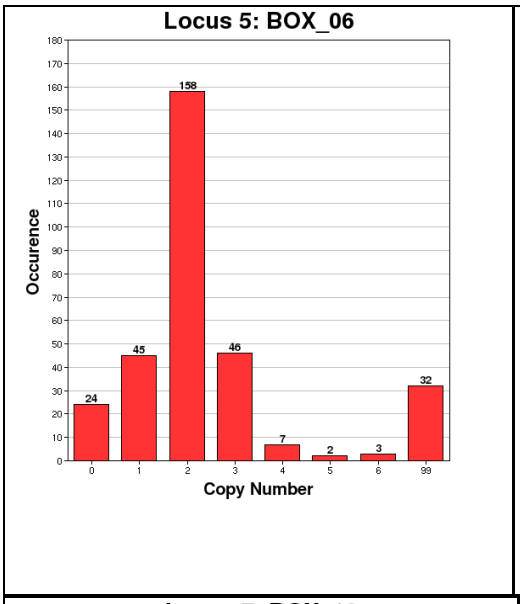
		Position in genome				
Primer	**Sequence**	**R6 (serotype 2)***	**Hungary19A-6 ** (19A)**	**G54 (19F)*****	**CGSP14 (14)^**	**TIGR4 (4)^^**
ms17f (B17) (other 16bp) 20-mer	TCGAAAATCTCTGCAAACCA	1783082-1783101 (40.1)	1974737-1974756 (40.1)	1812245-1812264 (40.1)	1945466-1945485 (40.1)	1901665-1901684 (40.1)
ms17r ***, ^, **, * 19-mer	CGGACTAGGCGGCTGATTA	1783248-1783234 (30.2)	1974993-1974979 (30.2)	1812365-1812351 (30.2)	1945676-1945662 (30.2)	1902235-1902217 (38.2)
Spneu17		**166**	**256**	**120**	**210**	**570**
ms19f 21-mer	TCGGGTGTAGTCGTGTTTACT	1925823-1925843 (42.1)	2131778-2131798 (42.1)	1960589-1960609 (42.1)	2093379-2093399 (42.1)	2046696-2046716 (42.1)
Ms19r 27-mer	AACTGATGTAGCTAAACCTAAAAAGAA	1926485-1926459 (54)	2132320-2132294 (54)	1961011-1960985 (54)	2093981-2093955 (54)	2047178-2047152 (54)
Spneu19		**662**	**542**	**422**	**602**	**482**
Ms25f (BOX-13) 20-mer	TCGCCTTTGCTAATCAAACA	101033- 101052 (40.1)	178817-178836 (40.1)	103496-103515 (40.1)	101226-101245 (40.1)	106484-106503 (40.1)
Ms25r 20-mer	GACTGGGTAACAATTCCATT	101458- 101439 (40.1)	179107-179088 (40.1)	103920-103901 (40.1)	101648-101629 (40.1)	106951-106932 (40.1)
Spneu25		**425**	**290**	**424**	**422**	**467**
Ms27f (B27)** 20-mer	TCAGGAACAGCTATTATCCC	257703- 257722 (40.1)	1042289-1042303 (30.2)	254840-254859 (40.1)	281039-281058 (40.1)	257247-257266 (40.1)
Ms27r** 20-mer	CCAACCTCCTTTTCGTTTCA	258049- 258030 (40.1)	174444-174455 (24.3)	255186-255167 (40.1)	281328-281309 (40.1)	257548-257529 (40.1)
Spneu27		**346**	**?**	**346**	**289**	**301**
Ms33f (BOX-04) (other 19bp) 20-mer	CAGCTGAACATGATGGCAAA	698391- 698410 (40.1)	822500-822519 (40.1)	698591-698610 (40.1)	730813-730832 (40.1)	743709-743728 (40.1)
Ms33r 21-mer	CATCACTTCTCCCTTCTAATC	698797- 698777 (42.1)	824374-824354 (42.1)	699087-699067 (42.1)	731264-731244 (42.1)	744250-744230 (42.1)
Spneu33		**406**	**1874**	**496**	**431**	**521**
Ms37f (BOX-12) (another 20bp) 20-mer	ATGCGCAAATCGATTAAGGA	1010066-1010047 (40.1) 1350250-1350269 (40.1)	1113368-1113349 (40.1) 1501188-1501207 (40.1)	1389989-1390008 (40.1)	1146548-1146567 (40.1) 1473097-1473116 (40.1)	1053517-1053498 (40.1) 1424329-1424348 (40.1)
Ms37r 20-mer	CGATGTGCTTCACTCACTCC	1350750-1350731 (40.1)	1501733-1501714 (40.1)	1390309-1390290 (40.1)	1473732-1473713 (40.1)	1424784-1424765 (40.1)
Spneu37		**500**	**545**	**320**	**635**	**455**
Ms39f (B39)*** 20-mer	CCTTGGAC(T)TACCACCTCGTT	1915770-1915789 (40.1)	2121623-2121642 (40.1)	1950488-1950507 (32.2)	2083269-2083288 (40.1)	2036537-2036556 (40.1)
Ms39r 20-mer	GCCGTGACAGACTTCTGGAA	1916044-1916025 (40.1)	2121987-2121968 (40.1)	1950807-1950788 (40.1)	2083588-2083569 (40.1)	2036904-2036885 (40.1)
Spneu39		**274**	**364**	**319**	**319**	**367**
BOX-10F2	GGAGCCGAGTAGGAGATTCTCAC	587893- 587915 (46.1)	705810-705832 (46.1)	592633-592655 (46.1)	619900-619922 (46.1)	631227-631249 (46.1)
BOX-10R2	TCGTAGGCTGCTACATTGACCAG	588403- 588381 (46.1)	706500-706478 (46.1)	593143-593121 (46.1)	620500-620478 (46.1)	631737-631715 (46.1)
BOX-10 (NEW)		**510**	**690**	**510**	**600**	**510**
BOX-12F2	GAGATTGCCCTTTTCATCTTCG	1350093-1350114 (44.1)	1501031-1501052 (44.1)	1389832-1389853 (44.1)	1472940-1472961 (44.1)	1424172-1424193 (44.1)
BOX-12R2	AGCAACCATTGAAACGCCTG	1350840-1350821 (40.1)	1501823-1501804 (40.1)	1390399-1390380 (40.1)	1473822-1473803 (40.1)	1424874-1424855 (40.1)

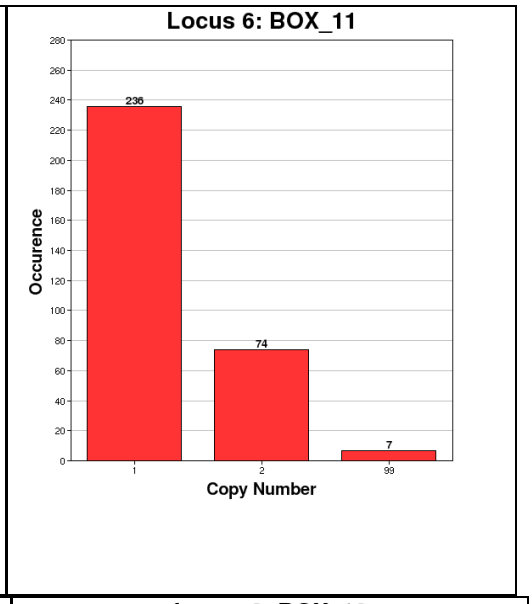
		Position in genome				
Primer	**Sequence**	**R6 (serotype 2)***	**Hungary19A-6 ** (19A)**	**G54 (19F)*****	**CGSP14 (14)^**	**TIGR4 (4)^^**
BOX-12 (NEW)		**747**	**792**	**567**	**882**	**702**
BOX-13F2	TCAAAAGATTGGAGAGTTCCGC	100977- 100998 (44.1)	178761-178782 (44.1)	103440-103461 (44.1)	101170-101191 (44.1)	106428-106449 (44.1)
BOX-13R2	GGATTTGGAGAGCAAGCAGATC	101549- 101528 (44.1)	179198-179177 (44.1)	104011-103990 (44.1)	101739-101718 (44.1)	107042-107021 (44.1)
BOX-13 (NEW)		**572**	**437**	**571**	**569**	**614**
Spneu19-F2	CACTCACCGTTAGCATTGACTCG	1925793-1925815 (46.1)	2131748-2131770 (46.1)	1960559-1960581 (46.1)	2093349-2093371 (46.1)	2046666-2046688 (46.1)
Spneu19-R2	TAATCAGGGAGTAGTTGGTTGGG	1926455-1926433 (46.1)	2132290-2132268 (46.1)	1960981-1960959 (46.1)	2093951-2093929 (46.1)	2047148-2047126 (46.1)
Spneu19 (NEW)		**662**	**520**	**422**	**602**	**482**

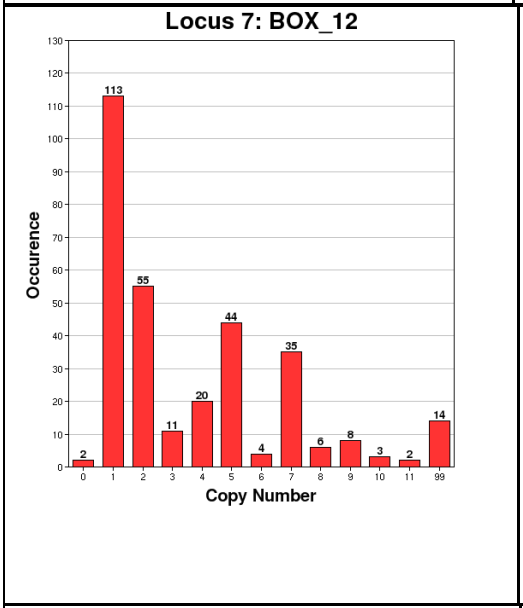
Appendix A3: Graphical representation of the diversity of MLVA genes in *S. pneumoniae*.

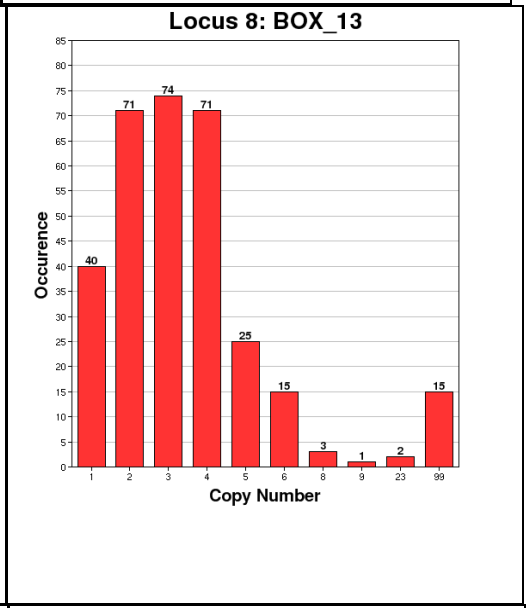
Number of occurrence for each locus is out of 317 *S. pneumoniae* isolates. Copy number represents the number of locus repeats for that allele (e.g. copy number 4 for Locus 1: BOX_01 indicates that BOX-01 has four repeats of *boxB* in the locus, and was observed in 14 *S. pneumoniae* isolates). A copy number of 99 indicate that the locus failed to amplify.

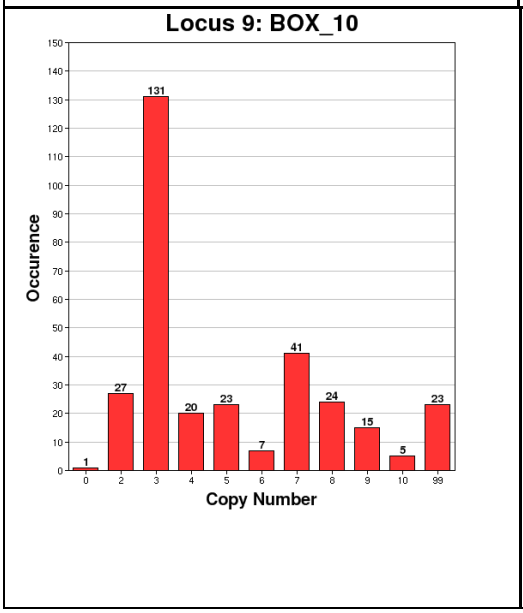


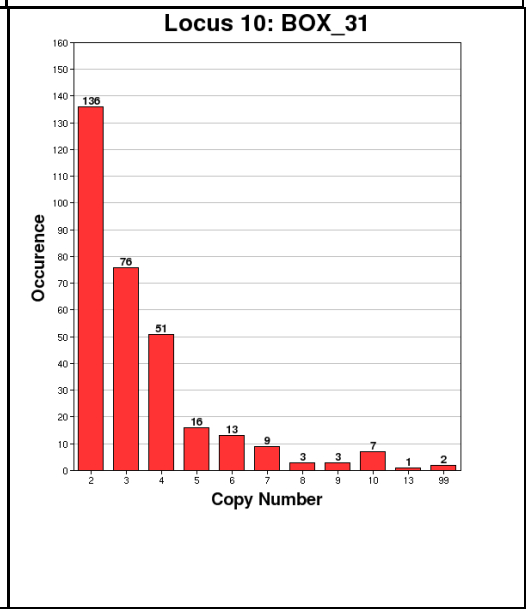

Locus 5: BOX_06
Occurence
Copy Number
24
45
158
46
7
2
3
32
0
1
2
3
4
5
6
99

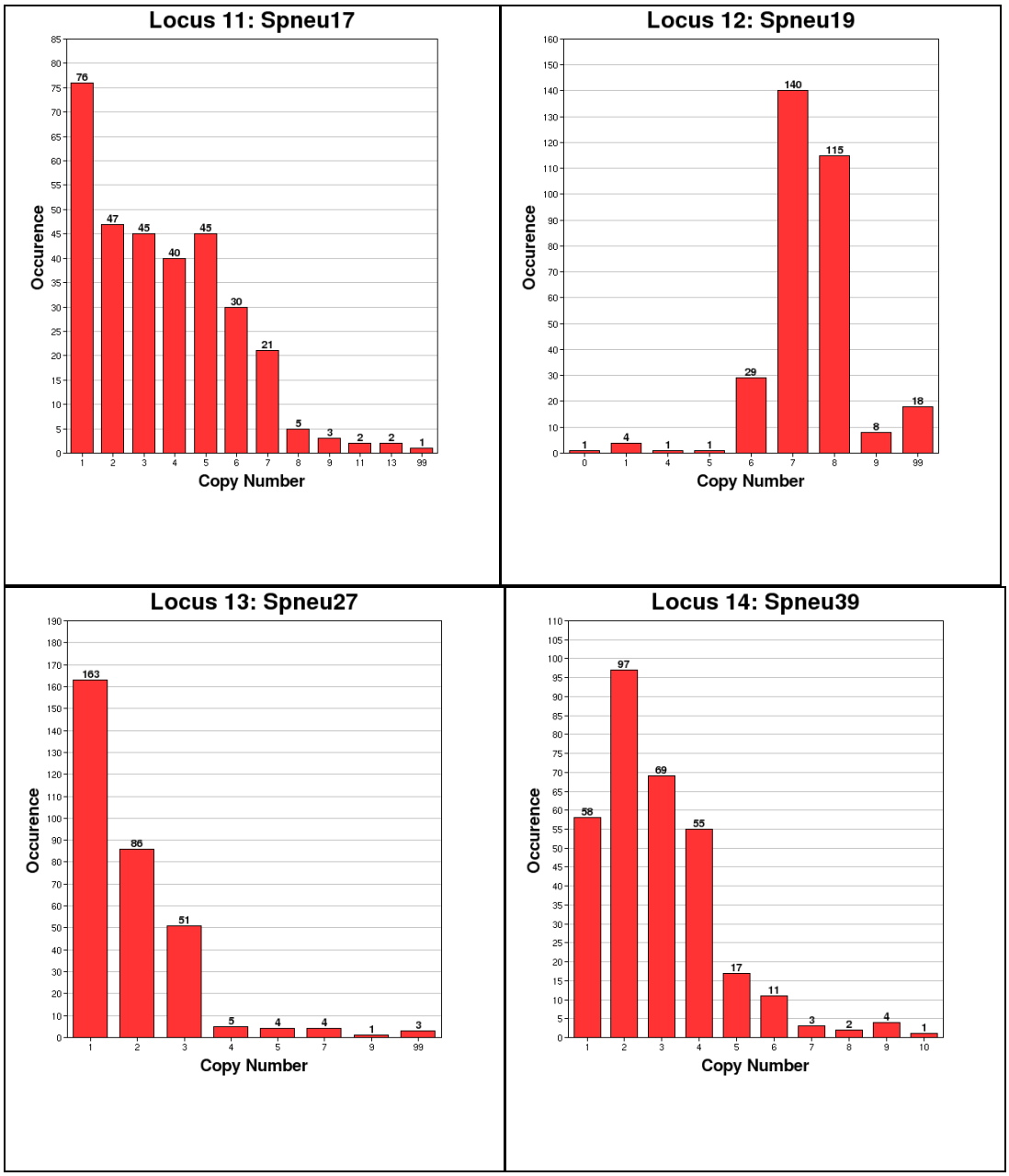

Locus 6: BOX_11
Occurence
Copy Number
236
74
7
1
2
99


Locus 7: BOX_12
Occurence
Copy Number
2
113
55
11
20
44
4
35
6
8
3
2
14
0
1
2
3
4
5
6
7
8
9
10
11
99


Locus 8: BOX_13
Occurence
Copy Number
40
71
74
71
25
15
3
1
2
15
1
2
3
4
5
6
8
9
23
99


Locus 9: BOX_10
Occurence
Copy Number
1
27
131
20
23
7
41
24
15
5
23
0
2
3
4
5
6
7
8
9
10
99


Locus 10: BOX_31
Occurence
Copy Number
136
76
51
16
13
9
3
3
7
1
2
2
3
4
5
6
7
8
9
10
13
99


Locus 11: Spneu17
Occurence
Copy Number
76
47
45
40
45
30
21
5
3
2
2
1
Locus 12: Spneu19
Occurence
Copy Number
1
4
1
1
29
140
115
8
18
Locus 13: Spneu27
Occurence
Copy Number
163
86
51
5
4
4
1
3
Locus 14: Spneu39
Occurence
Copy Number
58
97
69
55
17
11
3
2
4
1

Appendix A4: Comparison of various pneumococcal PCR-based serotyping methods.

Name	Methodology	Pros.	Cons.	Ref.
One-step multiplex PCR serotyping	Single amplification of *cpsA, wzy, wciP, wcwH*	100% sensitive and specific	Antiserum needed for serotype 6, 7, 9 and 18. Detects only 13vPCV serotypes.	Coskun-Ari *et al.,* 2012
Triplex real-time PCR serotyping	7 triplexed reaction with rtPCR; amplifies *wzy, wciP, wciN, wzx, wcwV*	High sensitivity and specificity. Tested 21 serotypes.	Expensive due to limited multiplexing	Pimenta *et al.,* 2013
Multiplex PCR serotyping	6 sequential multiplex PCR	Sequential PCR reduces time and cost; can be modified to suit various countries	Not all serotypes covered	Antonio *et al.,* 2009; Azzari *et al.,* 2010; Dias *et al.,* 2007; Iraurgui *et al.,* 2010; Jourdain *et al.,* 2011; Miernyk *et al.,* 2011; Njanpop Lafourcade *et al.,* 2010; Pai *et al.,* 2006; Saha *et al.,* 2008; Siira *et al.,* 2012
Microarray serotyping	274 oligonucleotide probes used for glycosyltransferase	Cost-effective and reliable. Tested 37 serotypes.	Time consuming due to number of probes	Tomita *et al.,* 2011
Multibead serotyping assay	Use of beads			Yu *et al.,* 2011
Real-time PCR serotyping	Use of rtPCR to amplify the capsule cassette	Directly applied to blood samples; lower limit of detection; tested 26/31 serotypes; no difference in sensitivity/specificity compared to sequential multiplex PCR		Azzari *et al.,* 2010; Marchese *et al.,* 2011; Tarrago *et al.,* 2008
Capsule Sequence Typing (CST)	Sequence 506bp of *wzh* and other specific genes	Tested serogroup 6 and 19; determine evolution	Only two serogroups tested; laborious to test routinely	Elberse *et al.,* 2011c
Real-time PCR serotyping	3 multiplex reactions using rtPCR	100% specificity; 17 serotypes tested	Fail to distinguish between some serotypes	Moore *et al.,* 2010

Name	Methodology	Pros.	Cons.	Ref.
Serotype-specific PCR	Two SNP-specific sense primers *wciP*584gS and *wciP*584aS	Distinguishes between serotype 6A, B and C	Only specific for serogroup 6	Jin *et al.,* 2009
Multiplex PCR-based Reverse Line Blot Hybridisation	20 sets of primers amplify in multiplex reactions and use of reverse line blot hybridisation	Type 90 serotypes	Labourious technique	Zhou *et al.,* 2007
PCR-RFLP serotyping	Amplify entire cps cassette and digested to produce serotype-specific patterns	Tested 46 serotypes	Difficult to analyse and compare between laboratories	Brugger *et al.,* 2009
Multiplex PCR (7 reactions)	7 multiplex reaction using PCR and gel electrophoresis	Distinguish between 9 serotypes targeted by 11vPCV; multiplexed	Only 9 serotypes detected, and only 65.9% (of 446 isolates) characterised to serogroup level	Brito *et al.,* 2003

Appendix A5: *In silico* analysis of 182 targets in the pneumococcal capsule cassette.

Each of 93 pneumococcal serotypes have a unique 'pseudoDNA' sequence consisting of T=gene present (highlighted yellow) and A=gene absent (not highlighted) from their capsule cassette (running left to right across table, genes are number 1 to 182). CPS loci are labelled across the top underneath gene number. Accession code is from the NCBI database for representative serotype used. Allele file code is designated against each pseudoDNA sequence and is used for the MinimumSNPs program. The whole pseudoDNA sequence will not fit across the page, therefore it has been broken up into sections (each section has a different border around the table).

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Accession	Serotype	allele file code	Tnp	wzg	wzh	wzd	wze	wchB	wchC	wchD	wzy	wzx	gla	ugd	rmlA	rmlC	rmlB	rmlD	wchA	wchF	wchG	wchH	wchI	glf	wchE	galU	pgm	HG261	wciI	wciJ	wciK	wciL	wciM	mnaA	fnlA	fnlB
CR931632	1	cps_>1	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	T	T	T	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T
CR931637	5	cps_>5	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	T	T
CR931638	6A	cps_>6	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM171374	6D	cps_>9	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931642	7C	cps_>12	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931645	9A	cps_>15	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931646	9L	cps_>16	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931647	9N	cps_>17	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931648	9V	cps_>18	T	T	T	T	T	A	A	A	T	T	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931649	10A	cps_>19	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931650	10B	cps_>20	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931651	10C	cps_>21	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931652	10F	cps_>22	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931653	11A	cps_>23	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931654	11B	cps_>24	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931656	11D	cps_>26	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931657	11F	cps_>27	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931658	12A	cps_>28	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	T	T	T
CR931659	12B	cps_>29	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	T	T	T

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Accession	Serotype	allele file code	Tnp	wzg	wzh	wzd	wze	wchB	wchC	wchD	wzy	wzx	gla	ugd	rmlA	rmlC	rmlB	rmlD	wchA	wchF	wchG	wchH	wchI	glf	wchE	galU	pgm	HG261	wcil	wciJ	wciK	wciL	wciM	mnaA	fnlA	fnlB
CR931660	12F	cps_>30	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	T	T	T
CR931661	13	cps_>31	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931663	15A	cps_>33	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931664	15B	cps_>34	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931665	15C	cps_>35	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931666	15F	cps_>36	T	T	T	T	T	A	A	A	T	T	A	A	A	A	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931667	16A	cps_>37	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931668	16F	cps_>38	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931670	17F	cps_>40	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931672	18B	cps_>42	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931673	18C	cps_>43	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931674	18F	cps_>44	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931675	19A	cps_>45	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931676	19B	cps_>46	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931677	19C	cps_>47	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	T	A	A
CR931678	19F	cps_>48	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931679	20A	cps_>49	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	T	A	A	A	A
JQ653093	20B	cps_>50	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	T	A	A	A	A
CR931680	21	cps_>51	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931683	23A	cps_>54	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931684	23B	cps_>55	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931685	23F	cps_>56	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931686	24A	cps_>57	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	T	T	T	T	T	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A
CR931690	25F	cps_>61	T	T	T	T	T	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A
CR931691	27	cps_>62	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931692	28A	cps_>63	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931693	28F	cps_>64	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931694	29	cps_>65	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931695	31	cps_>66	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931696	32A	cps_>67	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931697	32F	cps_>68	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931698	33A	cps_>69	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931699	33B	cps_>70	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931700	33C	cps_>71	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Accession	Serotype	allele file code	Tnp	wzg	wzh	wzd	wze	wchB	wchC	wchD	wzy	wzx	gla	ugd	rmlA	rmlC	rmlB	rmlD	wchA	wchF	wchG	wchH	wchI	glf	wchE	galU	pgm	HG261	wciI	wciJ	wciK	wciL	wciM	mnaA	fnlA	fnlB
CR931701	33D	cps_>72	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931702	33F	cps_>73	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931703	34	cps_>74	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931704	35A	cps_>75	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931705	35B	cps_>76	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931706	35C	cps_>77	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931707	35F	cps_>78	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931708	36	cps_>79	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	T	A	A
CR931709	37	cps_>80	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931710	38	cps_>81	T	T	T	T	T	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A
CR931711	39	cps_>82	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931712	40	cps_>83	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931714	41F	cps_>85	T	T	T	T	T	A	A	A	T	T	A	T	T	T	T	T	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931715	42	cps_>86	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	T	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931716	43	cps_>87	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931717	44	cps_>88	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	T	T	T
CR931718	45	cps_>89	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	A	A	A	A	A	T	A	A	A	A	T	T	A	T	A	A	T	T
CR931719	46	cps_>90	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	T	T	T
CR931720	47A	cps_>91	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931721	47F	cps_>92	T	T	T	T	T	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931722	48	cps_>93	T	T	T	T	T	A	A	A	T	T	A	A	T	T	T	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A

			35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Accession	Serotype	allele file code	fnlC	whaC	whaD	whaE	wciN	HG262	wciO	wciP	HG263	wcwA	wcwC	wcwD	HG140	wcwF	wcwG	wcwH	wcwI	wcwL	wcwK	wcxU	rbsF	wciR	wciS	wciT	HG265	HG266	wchO	wcjA	wcjB	wcjC	wcjD	wcjE	wcjG	wciB	wcrB
CR931632	1	cps_>1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931637	5	cps_>5	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931638	6A	cps_>6	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	A	A	A	A	T	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	A	A	A	A	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM17137	6D	cps_>9	A	A	A	A	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Accession	Serotype	allele file code	fnlC	whaC	whaD	whaE	wciN	HG262	wciO	wciP	HG263	wcwA	wcwC	wcwD	HG140	wcwF	wcwG	wcwH	wcwI	wcwL	wcwK	wcxU	rbsF	wciR	wciS	wciT	HG265	HG266	wchO	wcjA	wcjB	wcjC	wcjD	wcjE	wcjG	wciB	wcrB
CR931642	7C	cps_>12	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A
CR931645	9A	cps_>15	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A	A
CR931646	9L	cps_>16	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	T	A	A	A
CR931647	9N	cps_>17	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	T	A	A	A
CR931648	9V	cps_>18	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A	A
CR931649	10A	cps_>19	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T
CR931650	10B	cps_>20	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T
CR931651	10C	cps_>21	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T
CR931652	10F	cps_>22	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T
CR931653	11A	cps_>23	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931654	11B	cps_>24	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931656	11D	cps_>26	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931657	11F	cps_>27	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931658	12A	cps_>28	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931659	12B	cps_>29	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931660	12F	cps_>30	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931661	13	cps_>31	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931663	15A	cps_>33	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931664	15B	cps_>34	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931665	15C	cps_>35	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931666	15F	cps_>36	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931667	16A	cps_>37	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931668	16F	cps_>38	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931670	17F	cps_>40	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931672	18B	cps_>42	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931673	18C	cps_>43	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931674	18F	cps_>44	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931675	19A	cps_>45	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A
CR931676	19B	cps_>46	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T	A	A	A	A	A	A	A	A
CR931677	19C	cps_>47	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T	A	A	A	A	A	A	A	A
CR931678	19F	cps_>48	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A
CR931679	20A	cps_>49	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
JQ653093	20B	cps_>50	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931680	21	cps_>51	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Accession	Serotype	allele file code	fnlC	whaC	whaD	whaE	wciN	HG262	wciO	wciP	HG263	wcwA	wcwC	wcwD	HG140	wcwF	wcwG	wcwH	wcwI	wcwL	wcwK	wcxU	rbsF	wciR	wciS	wciT	HG265	HG266	wchO	wcjA	wcjB	wcjC	wcjD	wcjE	wcjG	wciB	wcrB
CR931683	23A	cps_>54	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931684	23B	cps_>55	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931685	23F	cps_>56	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931686	24A	cps_>57	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931690	25F	cps_>61	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931691	27	cps_>62	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931692	28A	cps_>63	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931693	28F	cps_>64	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931694	29	cps_>65	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931695	31	cps_>66	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A
CR931696	32A	cps_>67	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931697	32F	cps_>68	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931698	33A	cps_>69	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931699	33B	cps_>70	A	A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931700	33C	cps_>71	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931701	33D	cps_>72	A	A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931702	33F	cps_>73	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931703	34	cps_>74	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931704	35A	cps_>75	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931705	35B	cps_>76	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931706	35C	cps_>77	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931707	35F	cps_>78	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931708	36	cps_>79	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A
CR931709	37	cps_>80	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931710	38	cps_>81	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931711	39	cps_>82	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931712	40	cps_>83	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931714	41F	cps_>85	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931715	42	cps_>86	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931716	43	cps_>87	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931717	44	cps_>88	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931718	45	cps_>89	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931719	46	cps_>90	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931720	47A	cps_>91	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A
CR931721	47F	cps_>92	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A
CR931722	48	cps_>93	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
Accession	SeroType	allele file code	wcrC	wcrD	wclF	wclG	wcrH	wcrG	wchJ	wchK	wcyK	wcrL	wcwT	wcwU	gcT	wcwR	wcxB	wcxD	wcxE	wcxF	rnnaB	whaG	abp1	abp2	wchL	wchM	wchN	wclY	lrp	wclZ	wchX	gTp1
CR931632	1	cps_>1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931637	5	cps_>5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931638	6A	cps_>6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM17137	6D	cps_>9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931642	7C	cps_>12	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931645	9A	cps_>15	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931646	9L	cps_>16	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931647	9N	cps_>17	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931648	9V	cps_>18	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931649	10A	cps_>19	T	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931650	10B	cps_>20	T	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931651	10C	cps_>21	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931652	10F	cps_>22	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931653	11A	cps_>23	A	A	A	A	A	A	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931654	11B	cps_>24	A	A	A	A	A	A	T	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	A	A	A	A	A	A	T	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931656	11D	cps_>26	A	A	A	A	A	A	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931657	11F	cps_>27	A	A	A	A	A	A	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931658	12A	cps_>28	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931659	12B	cps_>29	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931660	12F	cps_>30	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931661	13	cps_>31	A	T	T	T	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A
CR931663	15A	cps_>33	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	T	T	T
CR931664	15B	cps_>34	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	T	T	T
CR931665	15C	cps_>35	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	T	T	T
CR931666	15F	cps_>36	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	T	T	T
CR931667	16A	cps_>37	A	A	A	T	A	A	T	T	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931668	16F	cps_>38	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931670	17F	cps_>40	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A
CR931672	18B	cps_>42	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A

			70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
Accession	SeroType	allele file code	wcrC	wcrD	wclF	wclG	wcrH	wcrG	wchJ	wchK	wcyK	wcrL	wcwT	wcwU	gcT	wcwR	wcxB	wcxD	wcxE	wcxF	mnaB	whaG	abp1	abp2	wchL	wchM	wchN	wclY	lrp	wclZ	wchX	gTp1
CR931673	18C	cps_>43	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A
CR931674	18F	cps_>44	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A
CR931675	19A	cps_>45	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931676	19B	cps_>46	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931677	19C	cps_>47	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931678	19F	cps_>48	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931679	20A	cps_>49	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
JQ653093	20B	cps_>50	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931680	21	cps_>51	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931683	23A	cps_>54	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
CR931684	23B	cps_>55	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
CR931685	23F	cps_>56	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
CR931686	24A	cps_>57	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931690	25F	cps_>61	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931691	27	cps_>62	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931692	28A	cps_>63	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931693	28F	cps_>64	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931694	29	cps_>65	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931695	31	cps_>66	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931696	32A	cps_>67	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931697	32F	cps_>68	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931698	33A	cps_>69	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931699	33B	cps_>70	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931700	33C	cps_>71	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931701	33D	cps_>72	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931702	33F	cps_>73	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931703	34	cps_>74	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931704	35A	cps_>75	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931705	35B	cps_>76	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931706	35C	cps_>77	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931707	35F	cps_>78	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931708	36	cps_>79	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931709	37	cps_>80	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931710	38	cps_>81	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931711	39	cps_>82	T	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931712	40	cps_>83	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
Accession	SeroType	allele file code	wcrC	wcrD	wciF	wciG	wcrH	wcrG	wchJ	wchK	wcyK	wcrL	wcwT	wcwU	gcT	wcwR	wcxB	wcxD	wcxE	wcxF	mnaB	whaG	abp1	abp2	wchL	wchM	wchN	wciY	lrp	wciZ	wchX	gTp1
CR931714	41F	cps_>85	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931715	42	cps_>86	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931716	43	cps_>87	T	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931717	44	cps_>88	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931718	45	cps_>89	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931719	46	cps_>90	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931720	47A	cps_>91	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931721	47F	cps_>92	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931722	48	cps_>93	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A

			100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129
Accession	SeroType	allele file code	gTp2	gTp3	wcxR	wcxS	wcxT	wciU	wcxM	wcxN	HG191	wcxP	wcxQ	wcxG	wcrT	wcrU	wcrV	wcrP	wcrQ	wcrR	wciV	wciW	wciX	HG94	wchP	wchQ	wchR	wchS	wchU	HG264	whaJ	wciD
CR931632	1	cps_>1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931637	5	cps_>5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931638	6A	cps_>6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM171374	6D	cps_>9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931642	7C	cps_>12	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931645	9A	cps_>15	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931646	9L	cps_>16	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931647	9N	cps_>17	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931648	9V	cps_>18	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931649	10A	cps_>19	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931650	10B	cps_>20	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931651	10C	cps_>21	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931652	10F	cps_>22	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931653	11A	cps_>23	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931654	11B	cps_>24	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129
Accession	SeroType	allele file code	gTp2	gTp3	wcxR	wcxS	wcxT	wciU	wcxM	wcxN	HG191	wcxP	wcxQ	wcxG	wcrT	wcrU	wcrV	wcrP	wcrQ	wcrR	wciV	wciW	wciX	HG94	wchP	wchQ	wchR	wchS	wchU	HG264	whaJ	wciD
CR931656	11D	cps_>26	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931657	11F	cps_>27	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931658	12A	cps_>28	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931659	12B	cps_>29	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931660	12F	cps_>30	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931661	13	cps_>31	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931663	15A	cps_>33	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931664	15B	cps_>34	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931665	15C	cps_>35	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931666	15F	cps_>36	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931667	16A	cps_>37	A	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931668	16F	cps_>38	A	A	A	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931670	17F	cps_>40	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	T	A	A	A	A	A	A	A	A
CR931672	18B	cps_>42	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A
CR931673	18C	cps_>43	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A
CR931674	18F	cps_>44	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A
CR931675	19A	cps_>45	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A
CR931676	19B	cps_>46	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A
CR931677	19C	cps_>47	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A
CR931678	19F	cps_>48	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A
CR931679	20A	cps_>49	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
JQ653093	20B	cps_>50	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T
CR931680	21	cps_>51	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931683	23A	cps_>54	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931684	23B	cps_>55	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931685	23F	cps_>56	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931686	24A	cps_>57	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931690	25F	cps_>61	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931691	27	cps_>62	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931692	28A	cps_>63	T	T	A	A	A	T	T	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931693	28F	cps_>64	T	T	A	A	A	T	T	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931694	29	cps_>65	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931695	31	cps_>66	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931696	32A	cps_>67	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A

			100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129
Accession	SeroType	allele file code	gTp2	gTp3	wcxR	wcxS	wcxT	wciU	wcxM	wcxN	HG191	wcxP	wcxQ	wcxG	wcrT	wcrU	wcrV	wcrP	wcrQ	wcrR	wciV	wciW	wciX	HG94	wchP	wchQ	wchR	wchS	wchU	HG264	whaJ	wciD
CR931697	32F	cps_>68	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A
CR931698	33A	cps_>69	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931699	33B	cps_>70	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931700	33C	cps_>71	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931701	33D	cps_>72	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931702	33F	cps_>73	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931703	34	cps_>74	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931704	35A	cps_>75	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931705	35B	cps_>76	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931706	35C	cps_>77	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931707	35F	cps_>78	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931708	36	cps_>79	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931709	37	cps_>80	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931710	38	cps_>81	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931711	39	cps_>82	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931712	40	cps_>83	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931714	41F	cps_>85	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A
CR931715	42	cps_>86	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931716	43	cps_>87	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931717	44	cps_>88	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931718	45	cps_>89	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931719	46	cps_>90	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931720	47A	cps_>91	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931721	47F	cps_>92	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931722	48	cps_>93	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
Accession	SeroType	allele file code	whaF	wcyT	wcyU	wcyO	wcwV	whaB	wcwX	wchV	wchW	HG267	wcxI	wcxJ	wcxK	HG268	HG269	wcyA	wcyB	wcyC	wcyD	wcyE	wcyF	whaK	whaL	wcyS	wcrN	HG270	wcjH	wcrJ	wcrM	wcrW
CR931632	1	cps_>1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
Accession	SeroType	allele file code	whaF	wcyT	wcyU	wcyO	wcwV	whaB	wcwX	wchV	wchW	HG267	wcxI	wcxJ	wcxK	HG268	HG269	wcyA	wcyB	wcyC	wcyD	wcyE	wcyF	whaK	whaL	wcyS	wcrN	HG270	wcjH	wcrJ	wcrM	wcrW
CR931637	5	cps_>5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931638	6A	cps_>6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM171374	6D	cps_>9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931642	7C	cps_>12	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931645	9A	cps_>15	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931646	9L	cps_>16	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931647	9N	cps_>17	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931648	9V	cps_>18	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931649	10A	cps_>19	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931650	10B	cps_>20	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931651	10C	cps_>21	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931652	10F	cps_>22	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931653	11A	cps_>23	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931654	11B	cps_>24	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931656	11D	cps_>26	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931657	11F	cps_>27	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931658	12A	cps_>28	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931659	12B	cps_>29	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931660	12F	cps_>30	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931661	13	cps_>31	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931663	15A	cps_>33	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931664	15B	cps_>34	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931665	15C	cps_>35	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931666	15F	cps_>36	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931667	16A	cps_>37	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931668	16F	cps_>38	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931670	17F	cps_>40	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931672	18B	cps_>42	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931673	18C	cps_>43	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931674	18F	cps_>44	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
Accession	SeroType	allele file code	whaF	wcyT	wcyU	wcyO	wcwV	whaB	wcwX	wchV	wchW	HG267	wcxI	wcxJ	wcxK	HG268	HG269	wcyA	wcyB	wcyC	wcyD	wcyE	wcyF	whaK	whaL	wcyS	wcrN	HG270	wcjH	wcrJ	wcrM	wcrW
CR931675	19A	cps_>45	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931676	19B	cps_>46	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931677	19C	cps_>47	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931678	19F	cps_>48	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931679	20A	cps_>49	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
JQ653093	20B	cps_>50	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931680	21	cps_>51	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	A	A	A	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931683	23A	cps_>54	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931684	23B	cps_>55	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931685	23F	cps_>56	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931686	24A	cps_>57	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	A	A	A	A	A	A	A	A	A	A	T	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	A	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A
CR931690	25F	cps_>61	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A
CR931691	27	cps_>62	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	A	A	A	A
CR931692	28A	cps_>63	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931693	28F	cps_>64	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931694	29	cps_>65	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A
CR931695	31	cps_>66	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931696	32A	cps_>67	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
CR931697	32F	cps_>68	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A
CR931698	33A	cps_>69	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931699	33B	cps_>70	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931700	33C	cps_>71	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931701	33D	cps_>72	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931702	33F	cps_>73	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931703	34	cps_>74	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931704	35A	cps_>75	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931705	35B	cps_>76	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A
CR931706	35C	cps_>77	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931707	35F	cps_>78	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931708	36	cps_>79	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931709	37	cps_>80	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931710	38	cps_>81	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	A	T	A	A	A	A	A	A	A	A	A
CR931711	39	cps_>82	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931712	40	cps_>83	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T

			130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
Accession	SeroType	allele file code	whaF	wcyT	wcyU	wcyO	wcwV	whaB	wcwX	wchV	wchW	HG267	wcxI	wcxJ	wcxK	HG268	HG269	wcyA	wcyB	wcyC	wcyD	wcyE	wcyF	whaK	whaL	wcyS	wcrN	HG270	wcjH	wcrJ	wcrM	wcrW
CR931714	41F	cps_>85	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T
CR931715	42	cps_>86	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
CR931716	43	cps_>87	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931717	44	cps_>88	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931718	45	cps_>89	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931719	46	cps_>90	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931720	47A	cps_>91	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931721	47F	cps_>92	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A
CR931722	48	cps_>93	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A

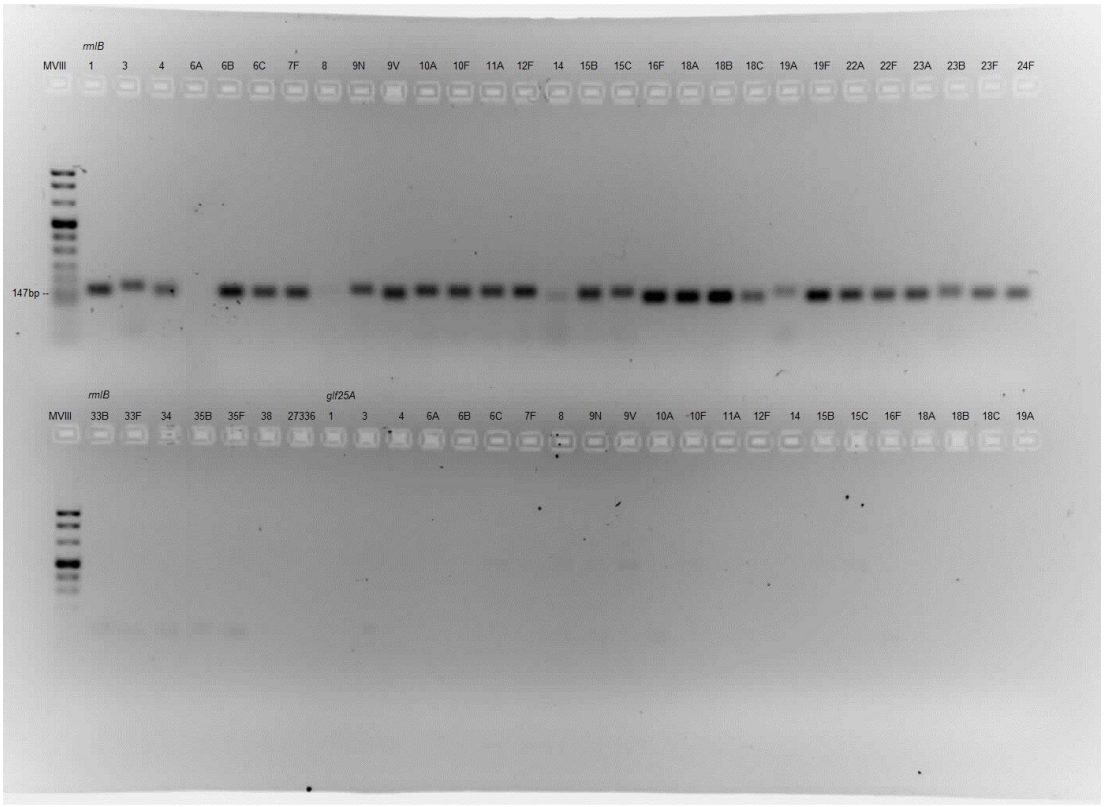
			160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
Accession	SeroType	allele file code	wcrX	wcyH	wcyI	HG272	HG271	wciC	wciE	wcrO	wcrI	wcrK	mnp1	mnp2	wcyV	whaI	wcyM	wcyN	wcyP	wcyQ	wcyR	HG273	wcwY	wciQ	whaM
CR931632	1	cps_>1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931633	2	cps_>2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931634	3	cps_>3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931635	4	cps_>4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931637	5	cps_>5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931638	6A	cps_>6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931639	6B	cps_>7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EF538714	6C	cps_>8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
HM171374	6D	cps_>9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931640	7A	cps_>10	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931641	7B	cps_>11	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931642	7C	cps_>12	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931643	7F	cps_>13	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931644	8	cps_>14	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A
CR931645	9A	cps_>15	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931646	9L	cps_>16	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931647	9N	cps_>17	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931648	9V	cps_>18	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931649	10A	cps_>19	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931650	10B	cps_>20	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

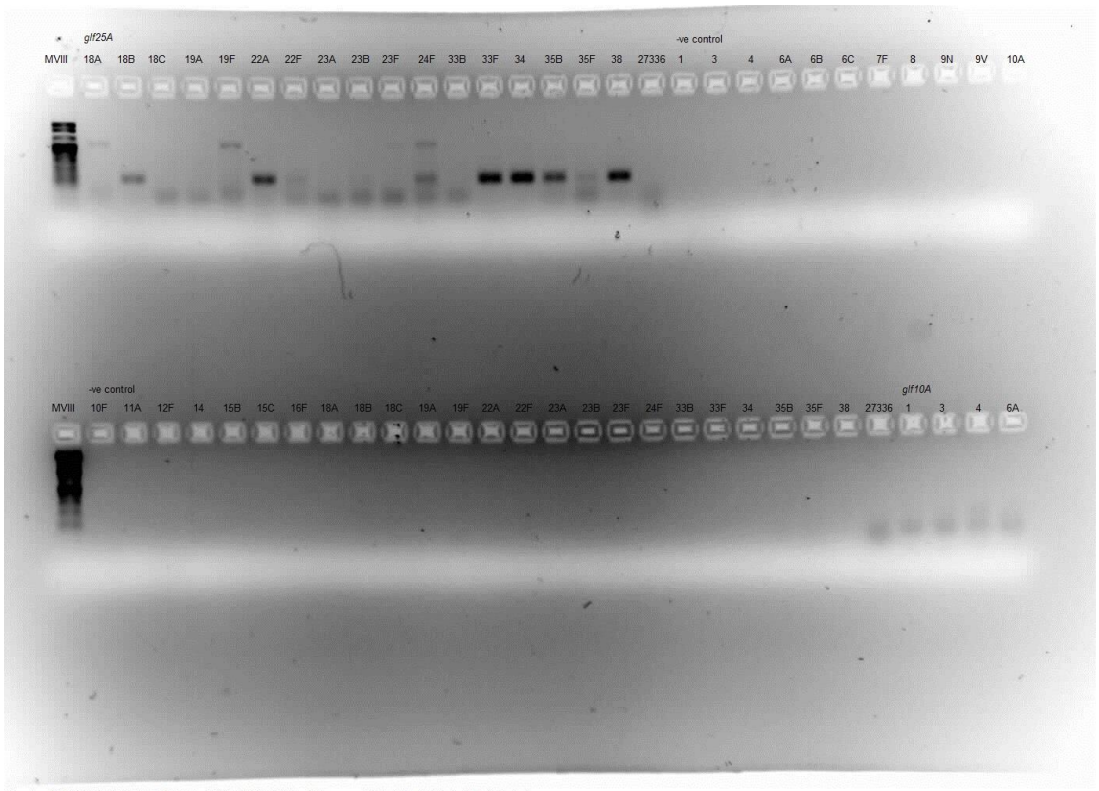
			160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
Accession	SeroType	allele file code	wcrX	wcyH	wcyI	HG272	HG271	wciC	wciE	wcrO	wcrI	wcrK	mnp1	mnp2	wcyV	whaI	wcyM	wcyN	wcyP	wcyQ	wcyR	HG273	wcwY	wciQ	whaM
CR931651	10C	cps_>21	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931652	10F	cps_>22	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931653	11A	cps_>23	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931654	11B	cps_>24	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931655	11C	cps_>25	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931656	11D	cps_>26	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931657	11F	cps_>27	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931658	12A	cps_>28	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931659	12B	cps_>29	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931660	12F	cps_>30	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931661	13	cps_>31	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931662	14	cps_>32	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931663	15A	cps_>33	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931664	15B	cps_>34	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931665	15C	cps_>35	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931666	15F	cps_>36	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931667	16A	cps_>37	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931668	16F	cps_>38	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931669	17A	cps_>39	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931670	17F	cps_>40	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931671	18A	cps_>41	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931672	18B	cps_>42	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931673	18C	cps_>43	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931674	18F	cps_>44	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931675	19A	cps_>45	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931676	19B	cps_>46	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931677	19C	cps_>47	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931678	19F	cps_>48	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931679	20A	cps_>49	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
JQ653093	20B	cps_>50	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931680	21	cps_>51	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931681	22A	cps_>52	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931682	22F	cps_>53	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931683	23A	cps_>54	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931684	23B	cps_>55	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931685	23F	cps_>56	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931686	24A	cps_>57	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931687	24B	cps_>58	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931688	24F	cps_>59	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931689	25A	cps_>60	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

			160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
Accession	SeroType	allele file code	wcrX	wcyH	wcyI	HG272	HG271	wciC	wciE	wcrO	wcrI	wcrK	mnp1	mnp2	wcyV	whaI	wcyM	wcyN	wcyP	wcyQ	wcyR	HG273	wcwY	wciQ	whaM
CR931690	25F	cps_>61	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931691	27	cps_>62	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931692	28A	cps_>63	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931693	28F	cps_>64	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931694	29	cps_>65	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931695	31	cps_>66	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931696	32A	cps_>67	A	T	T	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931697	32F	cps_>68	A	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931698	33A	cps_>69	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931699	33B	cps_>70	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931700	33C	cps_>71	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931701	33D	cps_>72	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931702	33F	cps_>73	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931703	34	cps_>74	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931704	35A	cps_>75	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931705	35B	cps_>76	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931706	35C	cps_>77	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931707	35F	cps_>78	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931708	36	cps_>79	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931709	37	cps_>80	A	A	A	A	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931710	38	cps_>81	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A
CR931711	39	cps_>82	A	A	A	A	A	A	T	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A
CR931712	40	cps_>83	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931713	41A	cps_>84	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931714	41F	cps_>85	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931715	42	cps_>86	A	A	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A
CR931716	43	cps_>87	A	A	A	A	A	A	T	A	A	A	A	A	A	T	T	T	T	A	A	A	A	A	A
CR931717	44	cps_>88	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931718	45	cps_>89	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A
CR931719	46	cps_>90	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CR931720	47A	cps_>91	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	A	A	A	A	A	A	T
CR931721	47F	cps_>92	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A
CR931722	48	cps_>93	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A

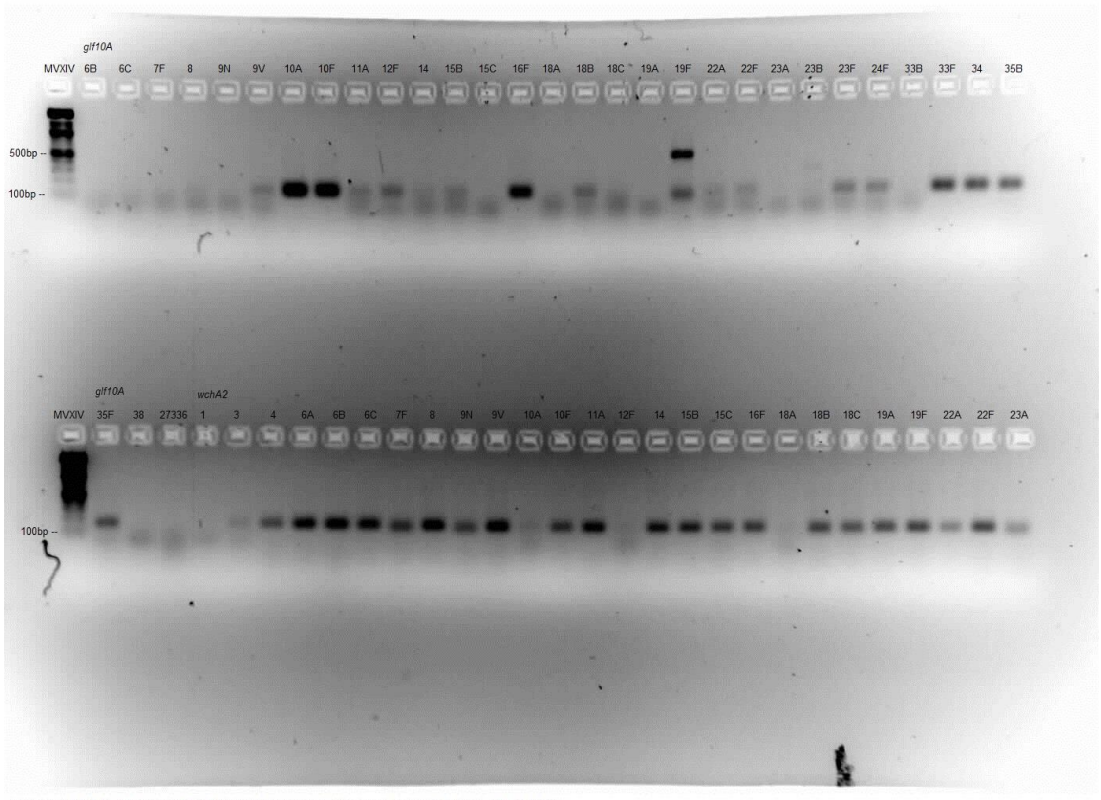
Appendix A6: Determination of presence/absence of 19 capsule targets in *S. pneumoniae* using conventional PCR and gel electrophoresis.

Marker ladder (MVIII) used in first column; targeted gene being amplified is written above the first serotype tested (i.e. serotype 1) and is therefore used for the pneumococcal isolates following until the next targeted gene is written; pneumococcal serotypes are written above each lane.

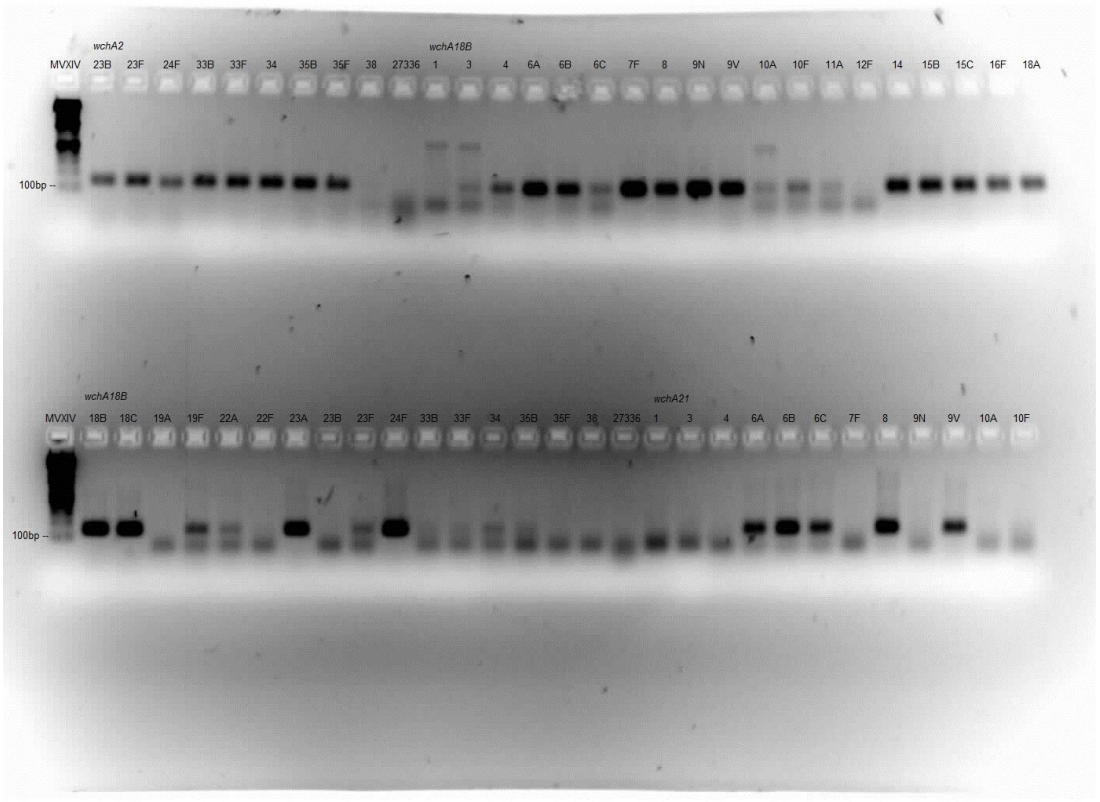




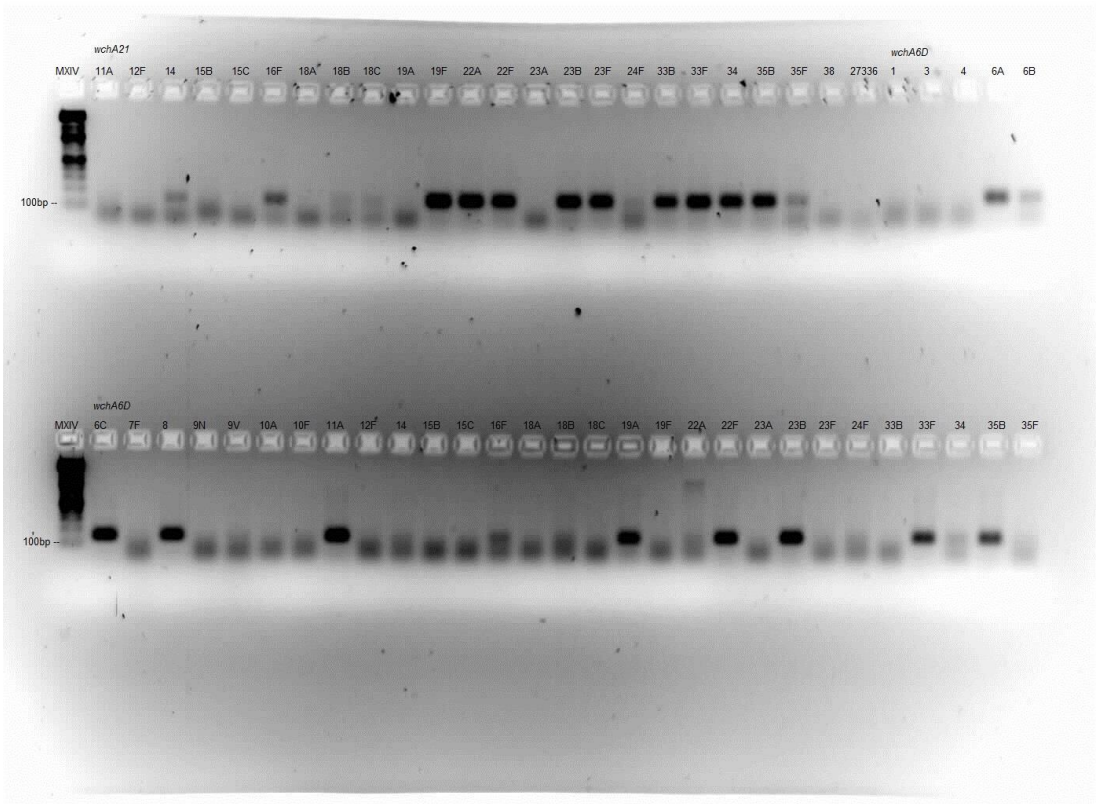
Tuesday, 13-05-14, 3:32:51PM B=2559 W=512 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 10uL Sybr Safe



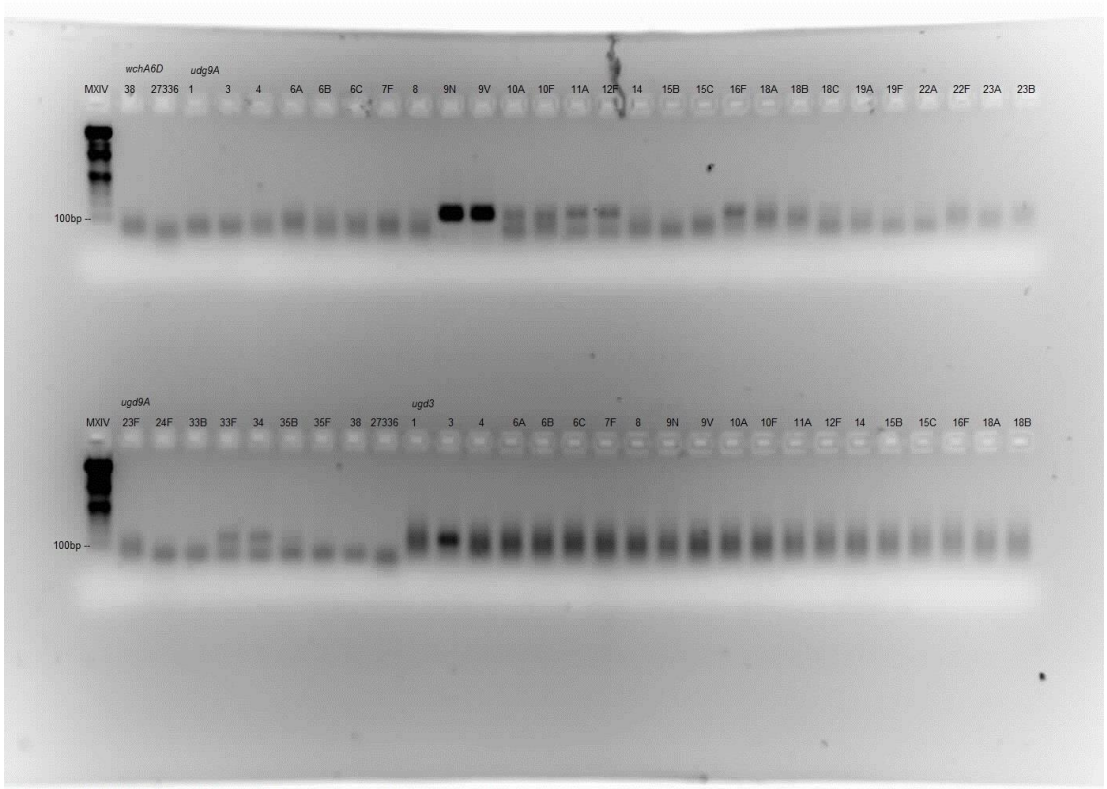
Wednesday, 14-05-14, 12:43:33PM B=2648 W=606 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe



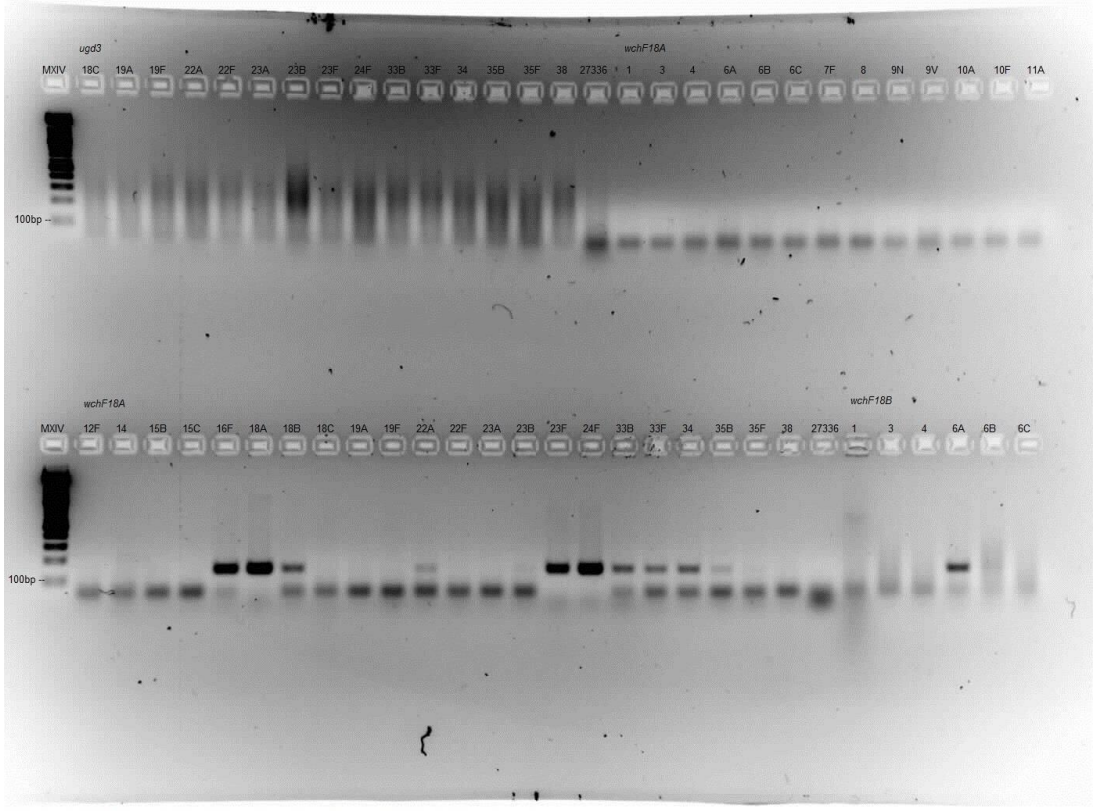
Thursday, 15-05-14, 10:38:52AM B=2650 W=607 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe



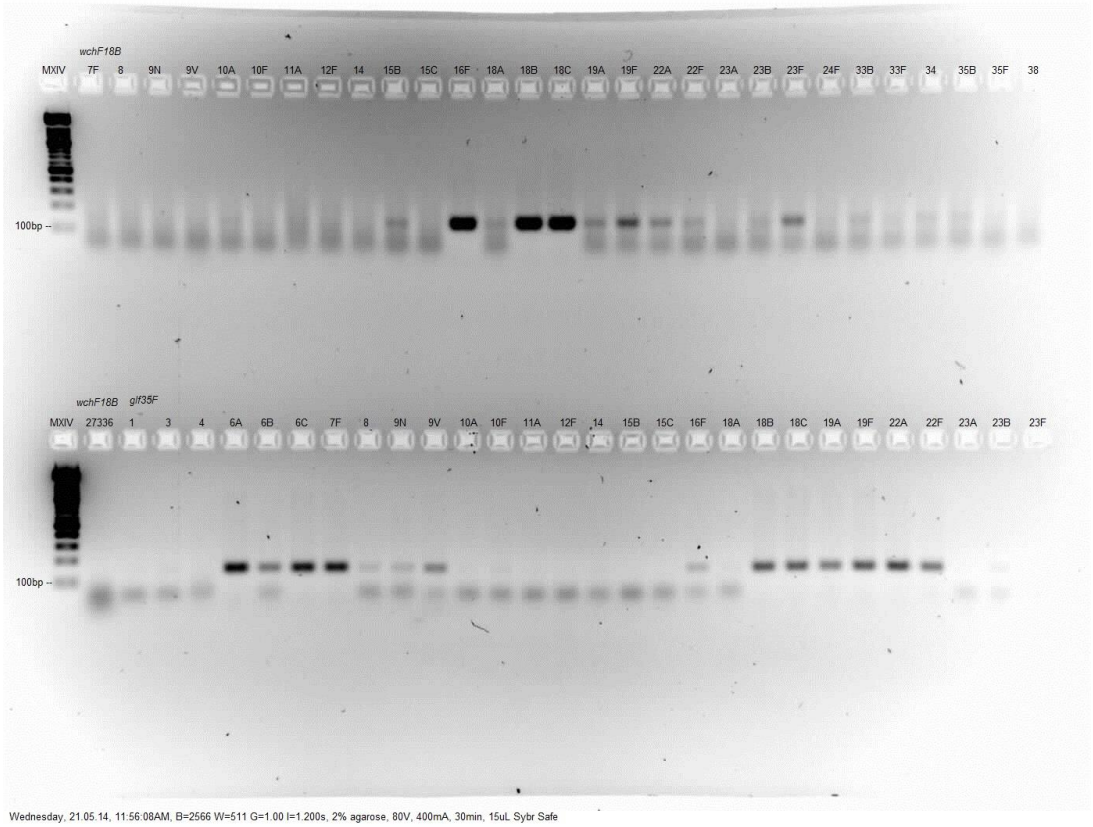
Thursday, 15-05-14, 12:05:06PM, B=2714 W=671 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe

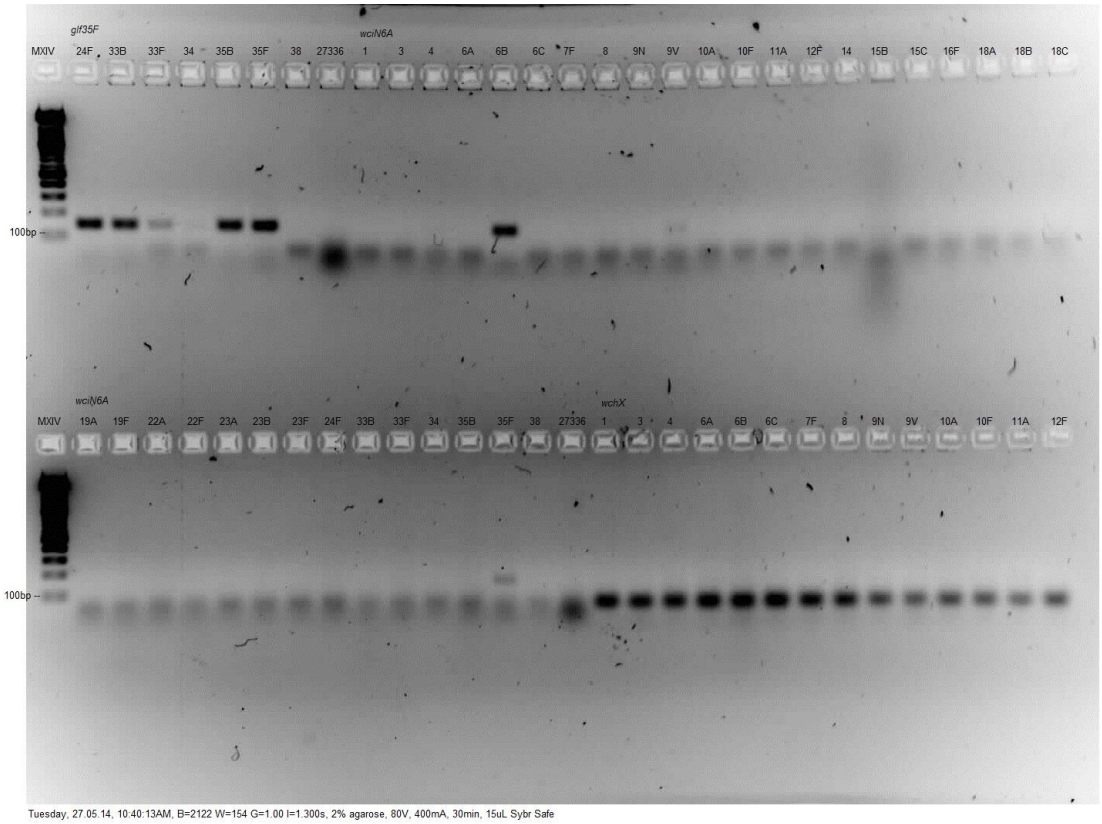


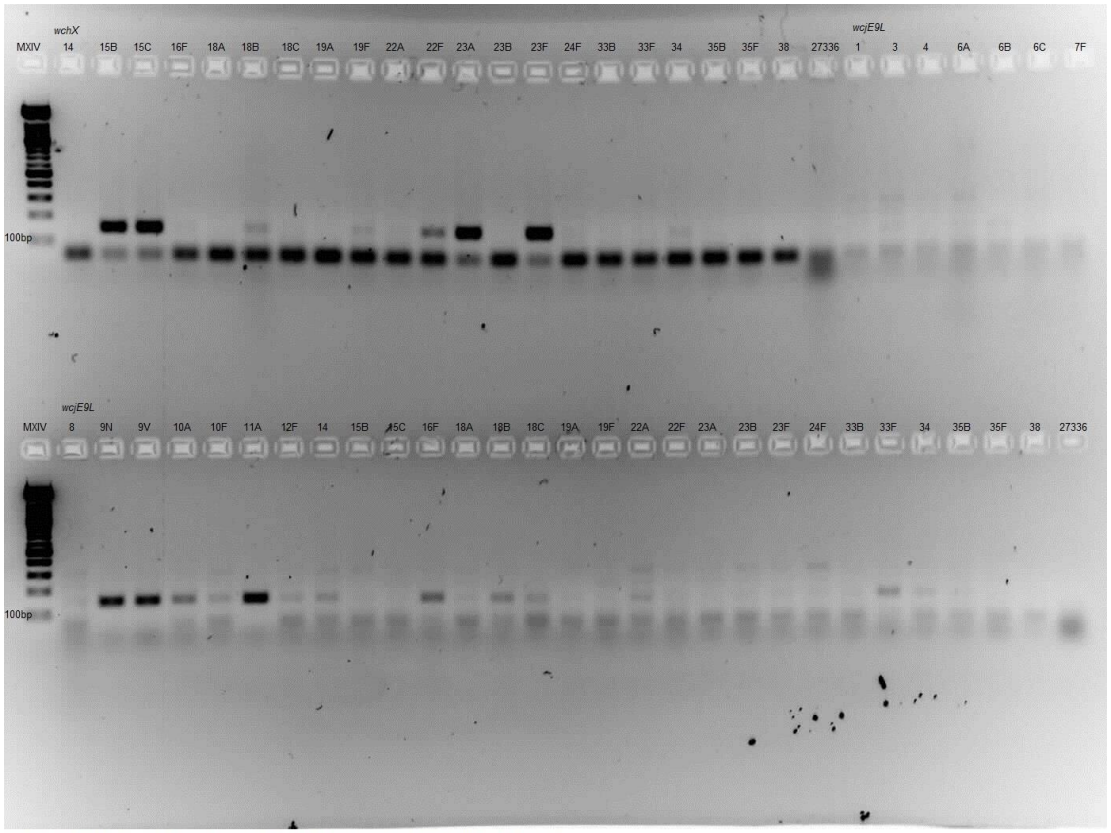
Thursday, 15.05.14, 1:01:40PM, B=3712 W=180 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe

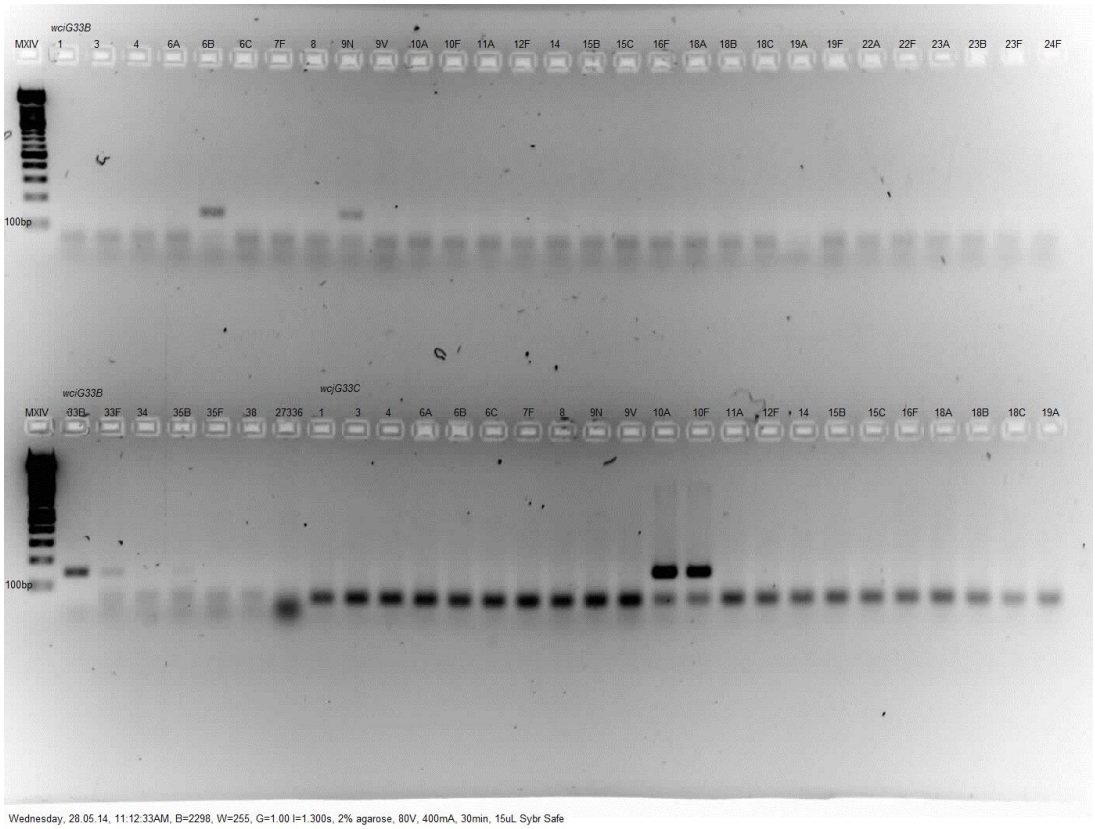


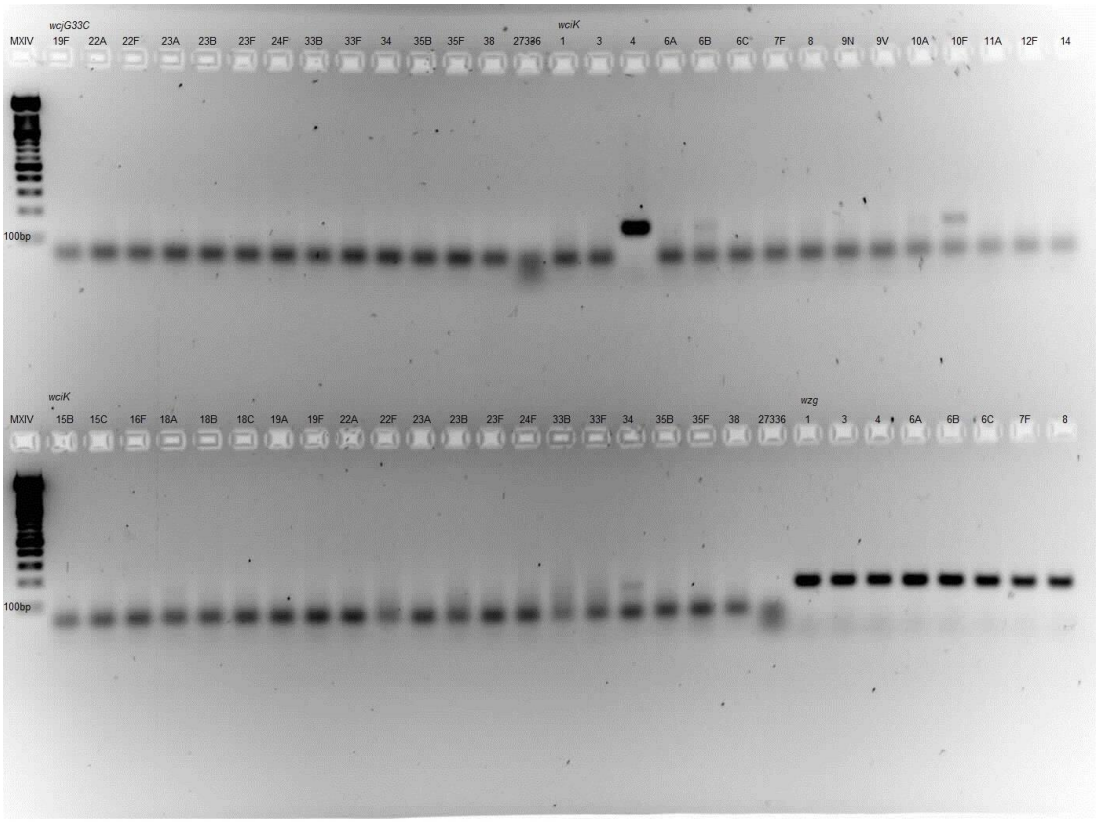
Wednesday, 21.05.14, 10:26:42AM, B=2391 W=351 G=1.00 I=1.300s, 2% agarose, 80V 400mA, 30min, 15uL Sybr Safe


wchF18B
MXIV 7F 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C 19A 19F 22A 22F 23A 23B 23F 24F 33B 33F 34 35B 35F 38
100bp --
wchF18B glf35F
MXIV 27336 1 3 4 6A 6B 6C 7F 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C 19A 19F 22A 22F 23A 23B 23F
100bp --
Wednesday, 21.05.14, 11:56:08AM, B=2566 W=511 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe

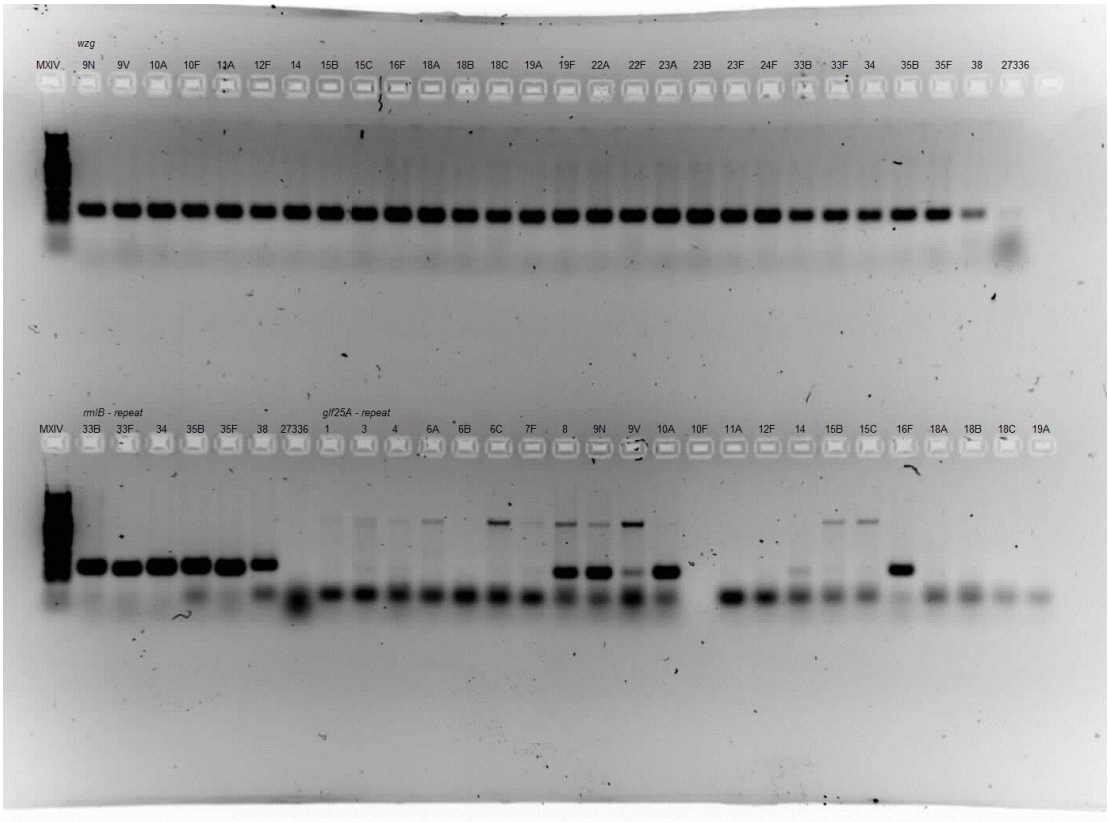

glf35F wciN6A
MXIV 24F 33B 33F 34 35B 35F 38 27336 1 3 4 6A 6B 6C 7F 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C
100bp --
wciN6A wchX
MXIV 19A 19F 22A 22F 23A 23B 23F 24F 33B 33F 34 35B 35F 38 27336 1 3 4 6A 6B 6C 7F 8 9N 9V 10A 10F 11A 12F
100bp --
Tuesday, 27.05.14, 10:40:13AM, B=2122 W=154 G=1.00 I=1.300s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe


wchX
wcjE9L
MXIV 14 15B 15C 16F 18A 18B 18C 19A 19F 22A 22F 23A 23B 23F 24F 33B 33F 34 35B 35F 38 27336 1 3 4 6A 6B 6C 7F
100bp
wcjE9L
MXIV 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C 19A 19F 22A 22F 23A 23B 23F 24F 33B 33F 34 35B 35F 38 27336
100bp
Wednesday, 28.05.14, 10:11:27AM, B=2113 W=95 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe


wciG33B
MXIV 1 3 4 6A 6B 6C 7F 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C 19A 19F 22A 22F 23A 23B 23F 24F
100bp
wciG33B
wcjG33C
MXIV 33B 33F 34 35B 35F 38 27336 1 3 4 6A 6B 6C 7F 8 9N 9V 10A 10F 11A 12F 14 15B 15C 16F 18A 18B 18C 19A
100bp
Wednesday, 28.05.14, 11:12:33AM, B=2298, W=255, G=1.00 I=1.300s, 2% agarose, 80V, 400mA, 30min, 15uL Sybr Safe



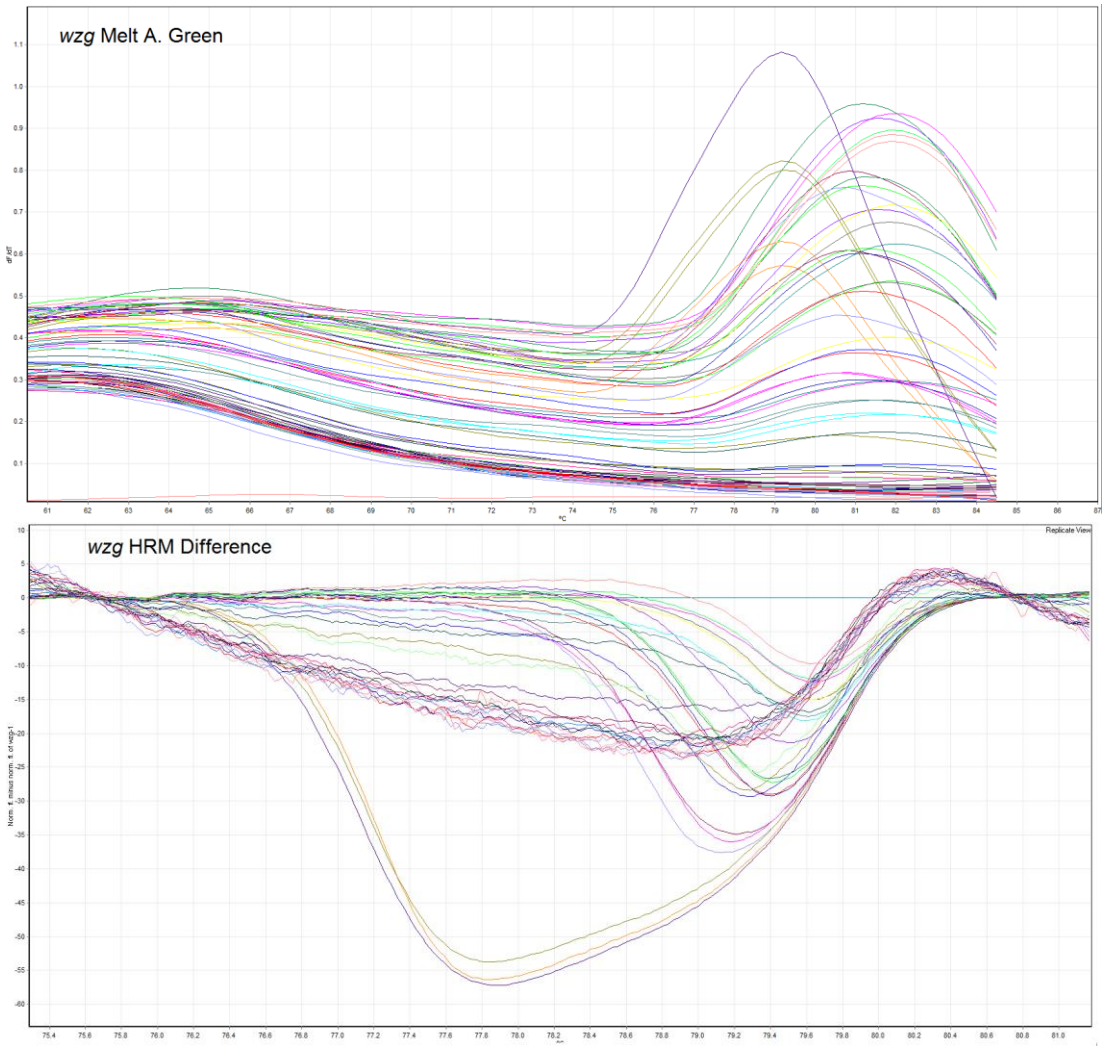
Wednesday, 28.05.14, 12:36:49PM, B=2291 W=255 G=1.00 I=1.200s, 2% agarose, 80V, 400mA, 30min 15uL Sybr Safe

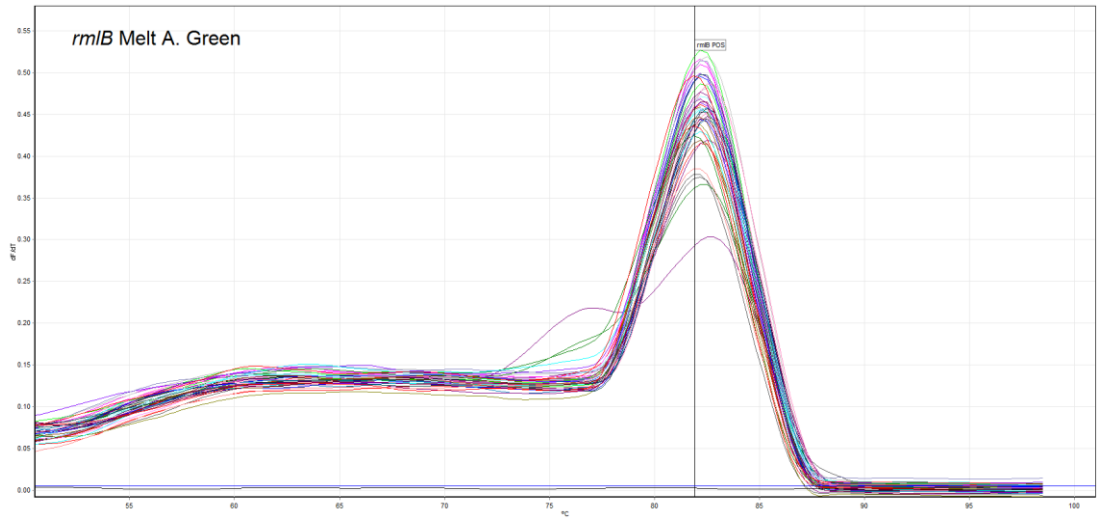


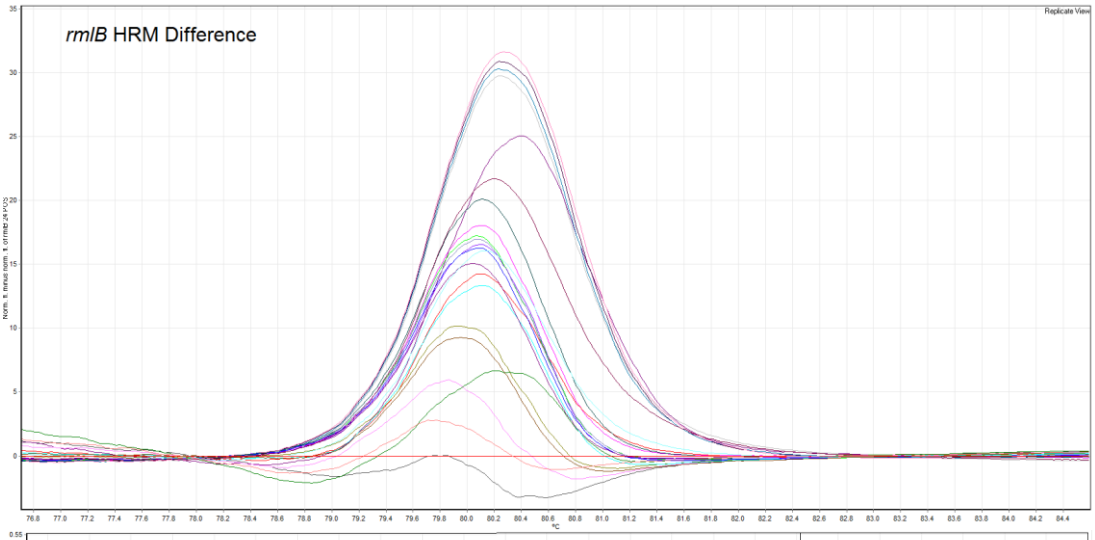
Monday, 2.06.14, 1:39:22PM, B=2143 W=96 G=1.00 I=1.200s, 80V, 400mA, 30min, 2% agarose, 15uL Sybr Safe

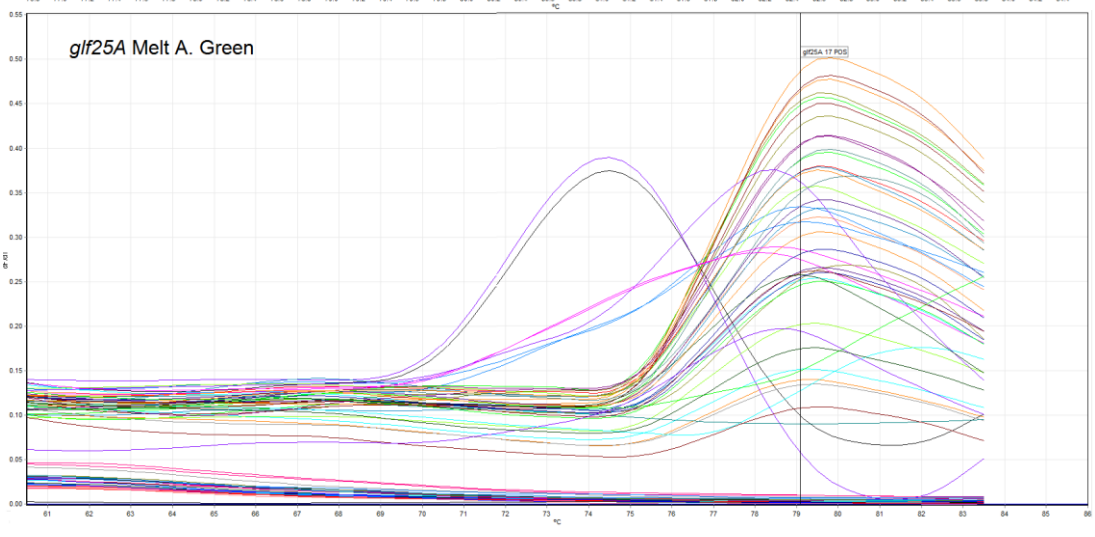
Appendix A7: Real-time PCR analysis of 19 primer targets in the *S. pneumoniae* capsule cassette.

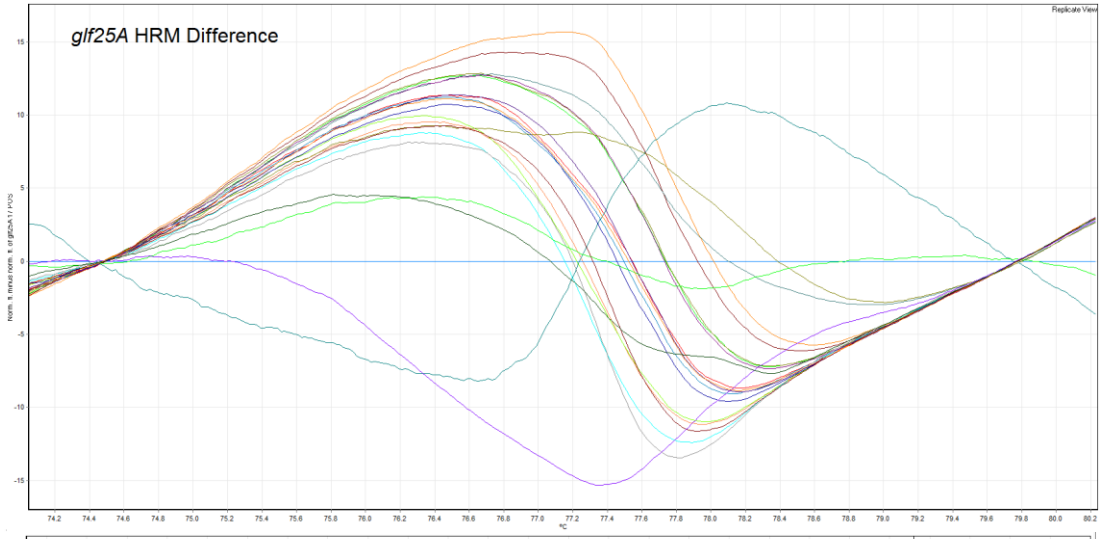
Melt A. Green heat curve and High Resolution Melt (HRM) Difference curve is presented for each locus. Melt A. Green heat curves contain a positive control strain, and peaks indicate a positive amplification of the loci. HRM Difference curve is based on the positive control, and a sample that is more than ±5 units from the positive control (set as the base line '0') indicates that the locus is significantly different than the positive control. HRM Difference samples have replicates combined for ease of analysis.

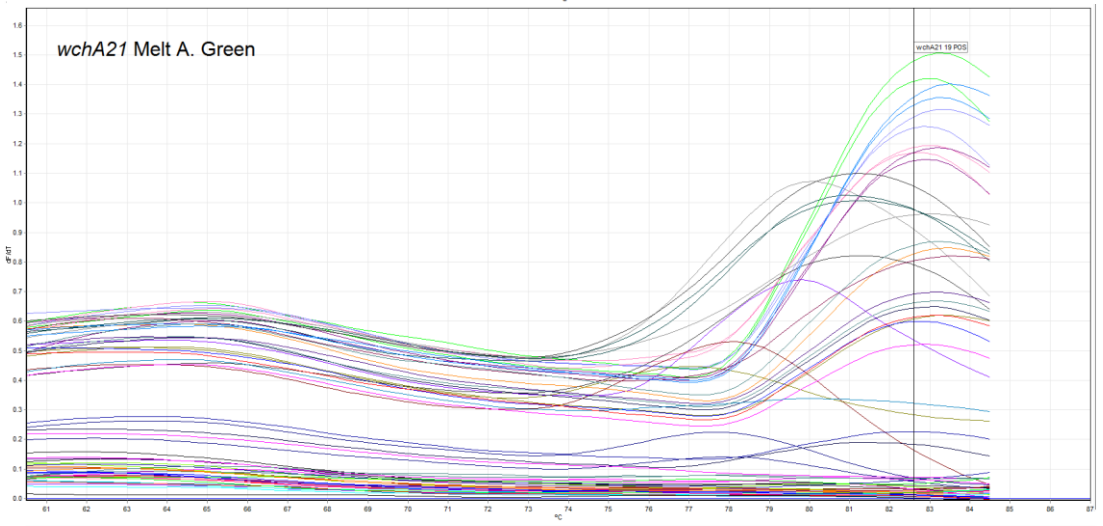


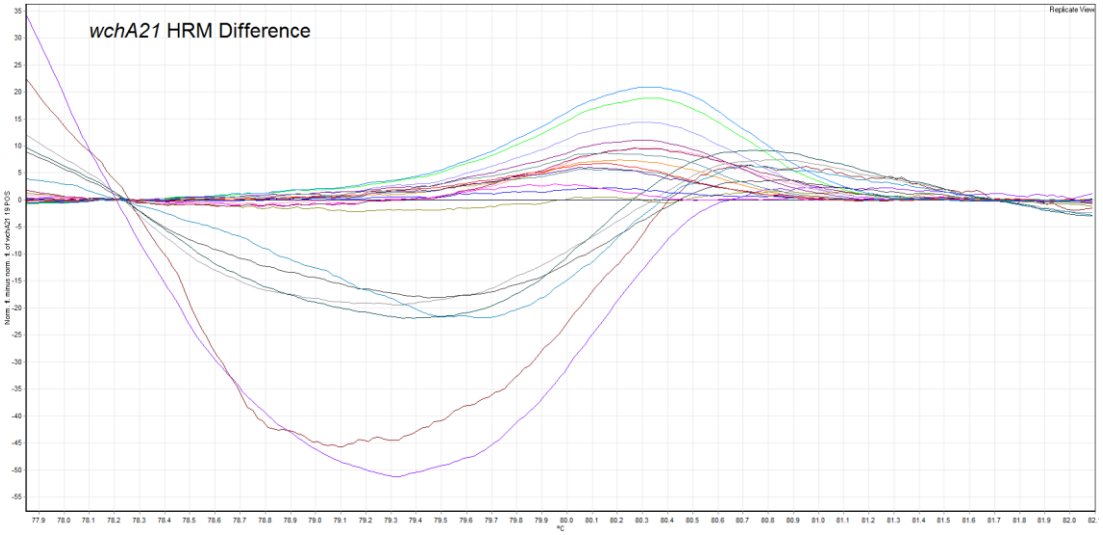

rmlB Melt A. Green
rmlB POS
-dF/dT
°C

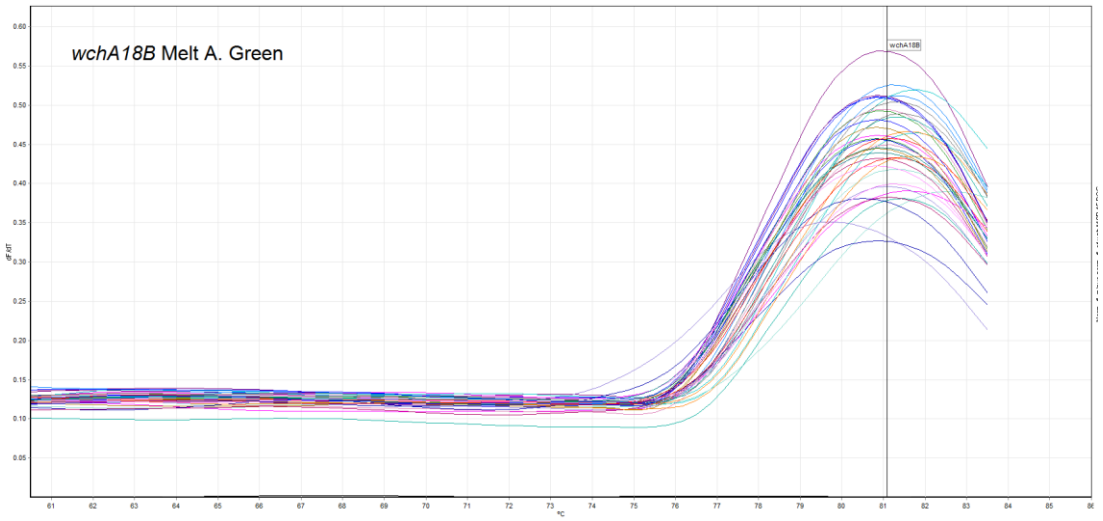

rmlB HRM Difference
Replicate View
°C

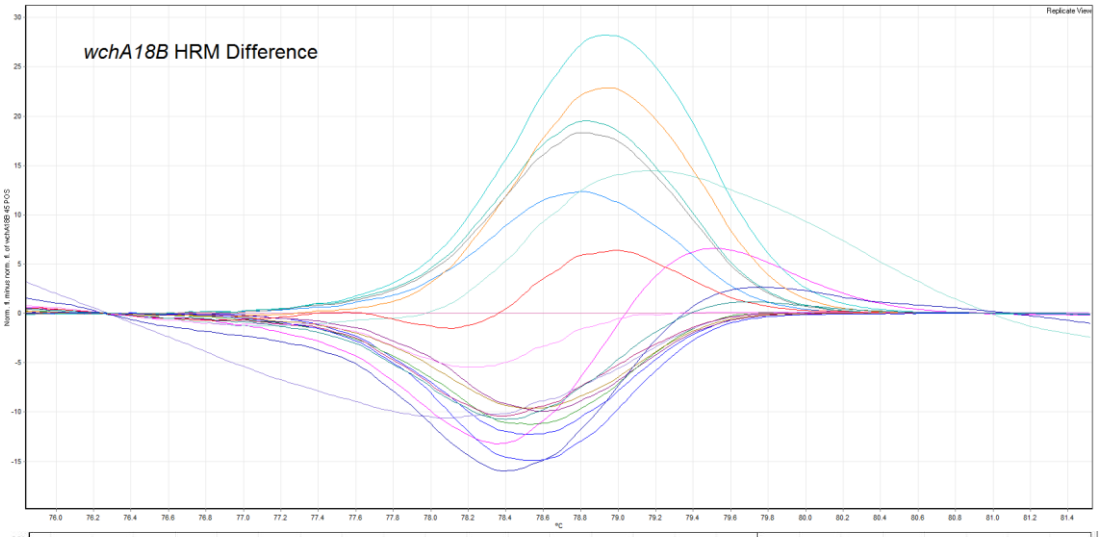

glf25A Melt A. Green
glf25A 17 POS
-dF/dT
°C

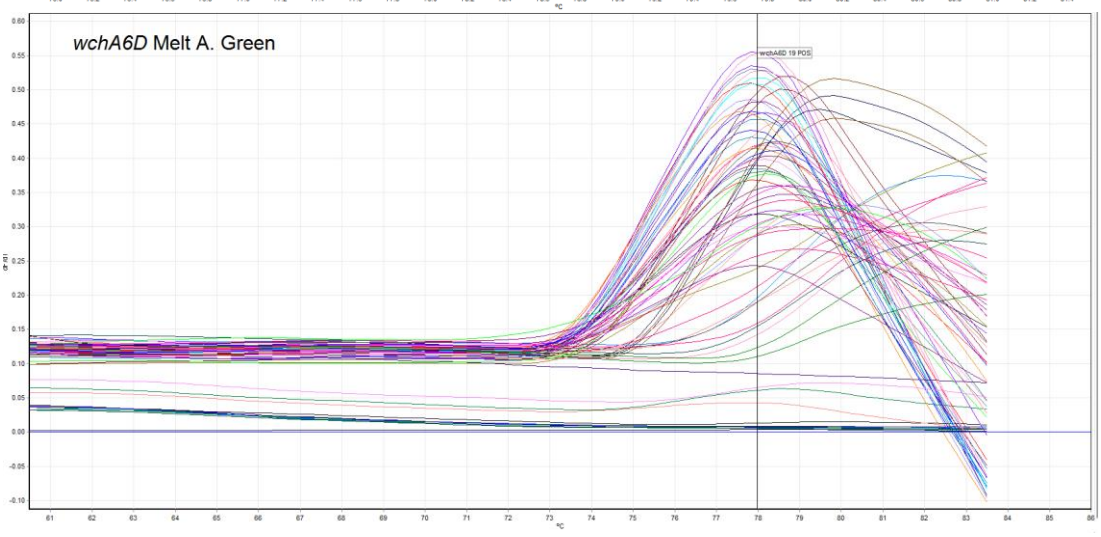

glf25A HRM Difference
Replicate View

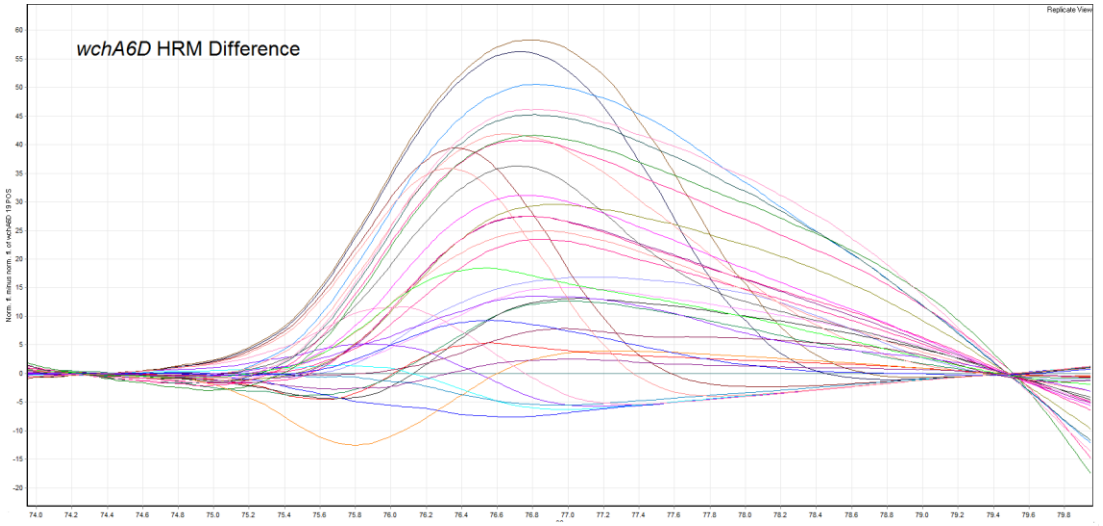

wchA21 Melt A. Green
wchA21 19 POS

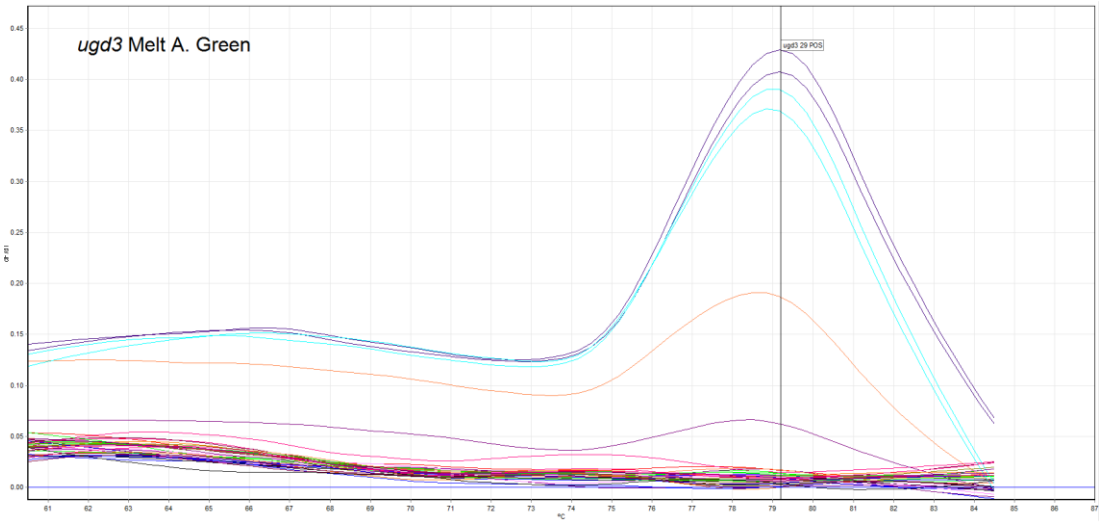

wchA21 HRM Difference
Replicate View

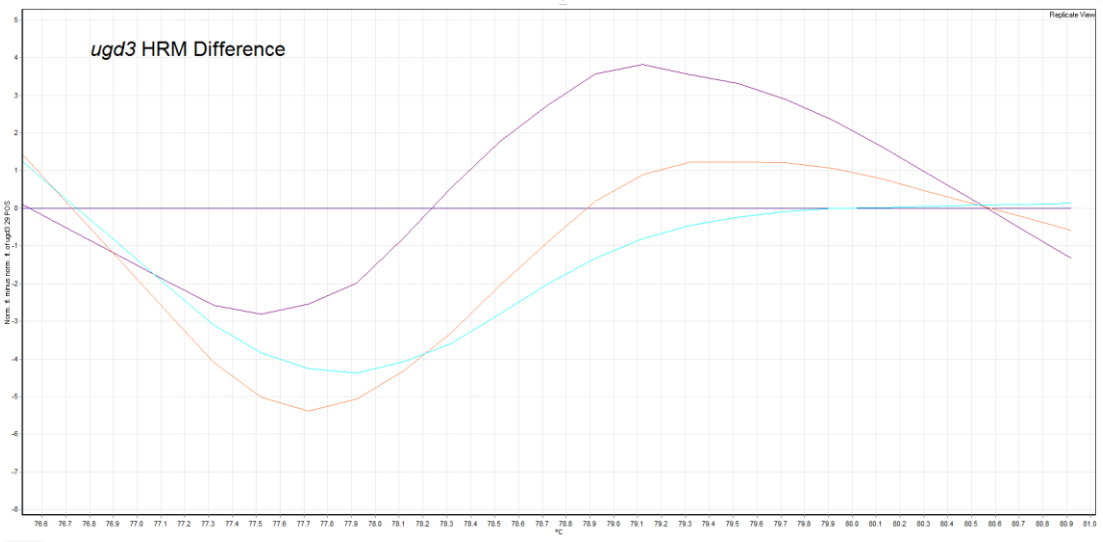

wchA18B Melt A. Green

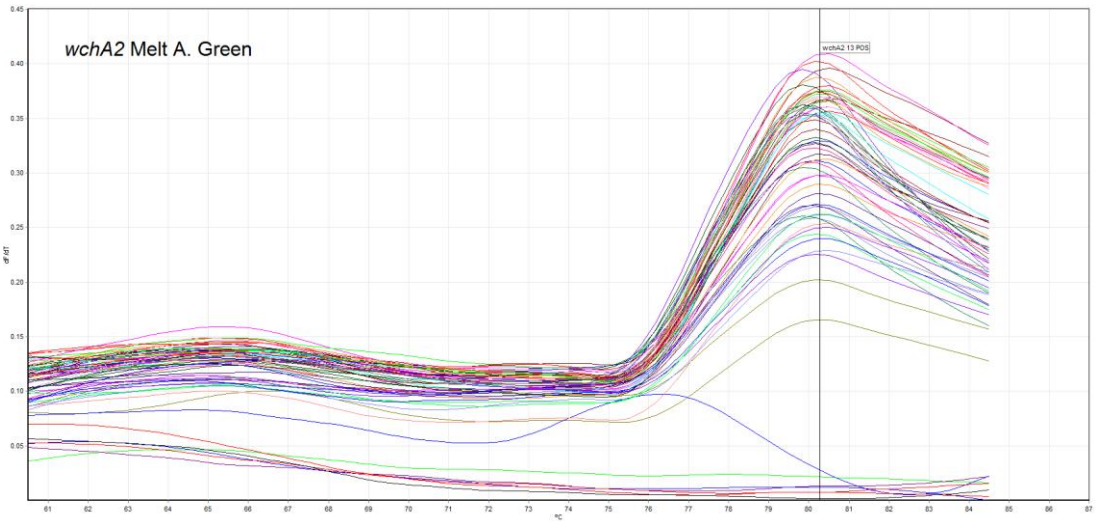

wchA18B HRM Difference

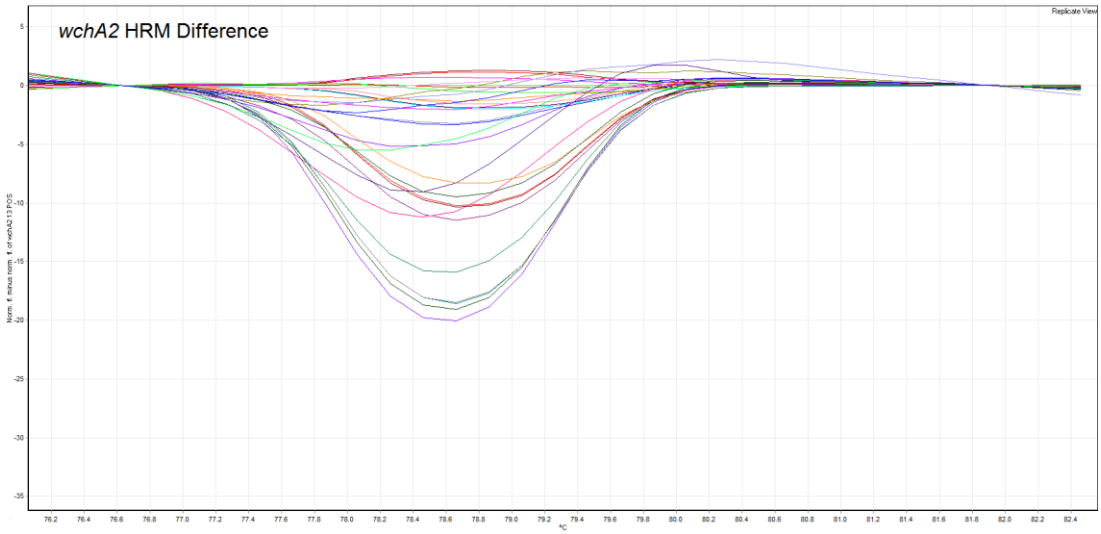

wchA6D Melt A. Green

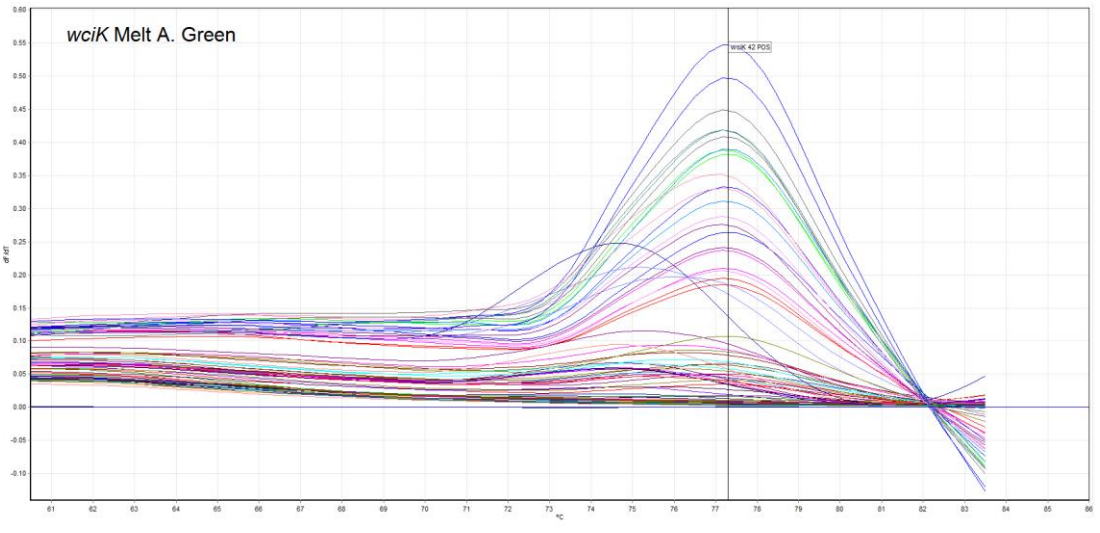

wchA6D HRM Difference
Replicate View

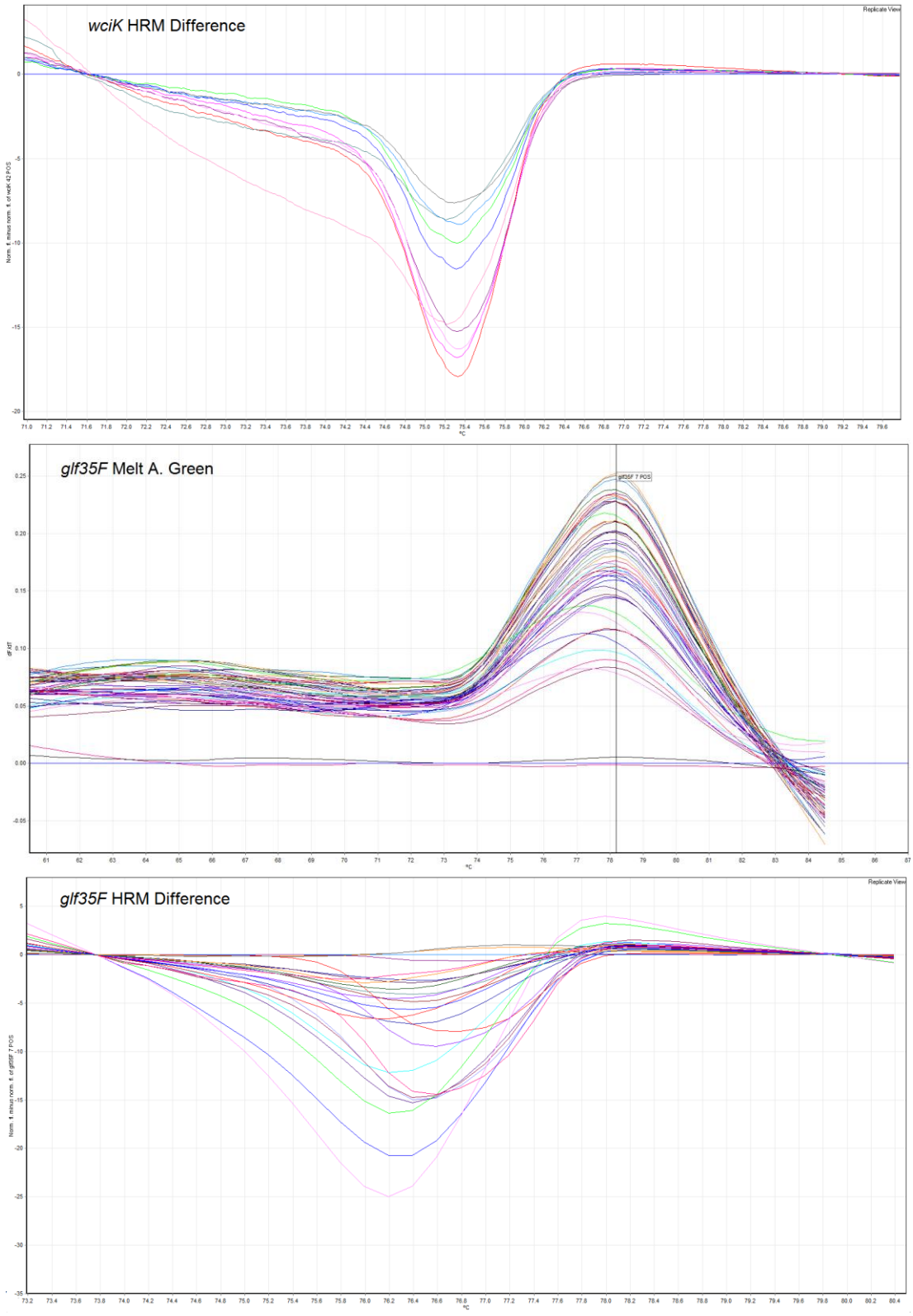

ugd3 Melt A. Green
ugd3 29 POS

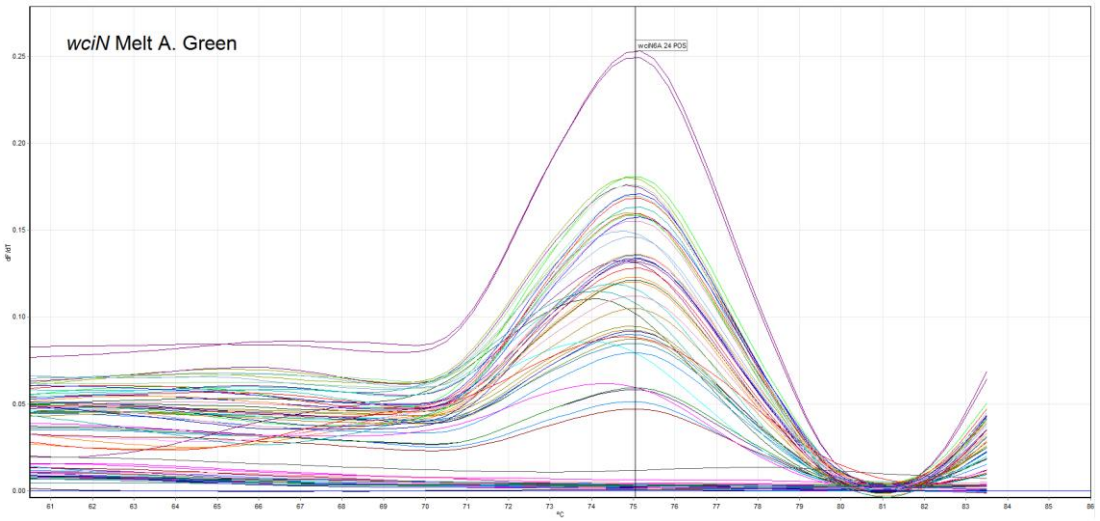

ugd3 HRM Difference
Replicate View

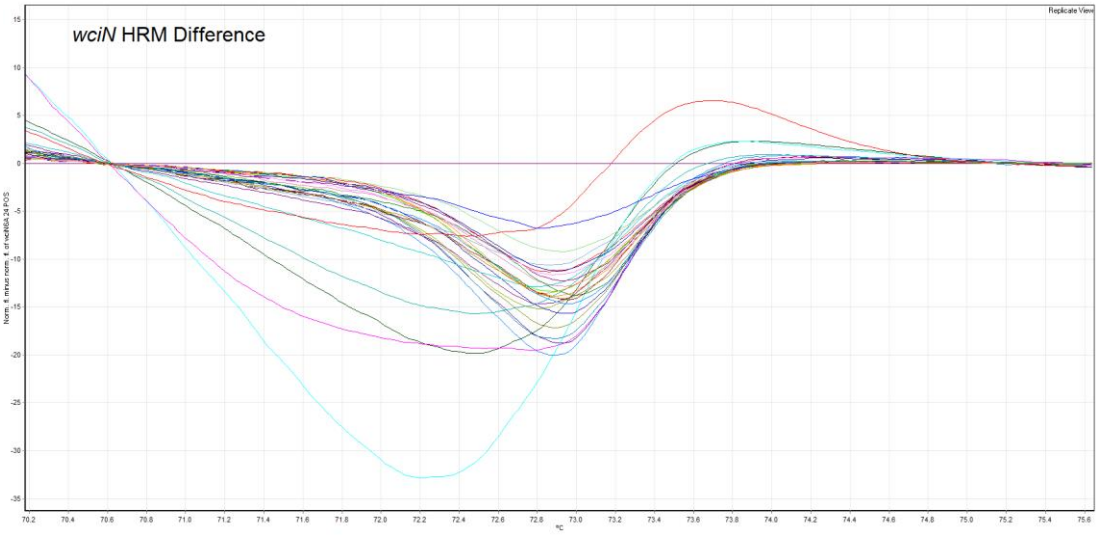

wchA2 Melt A. Green
wchA2 13 POS
-d'F/dT
°C

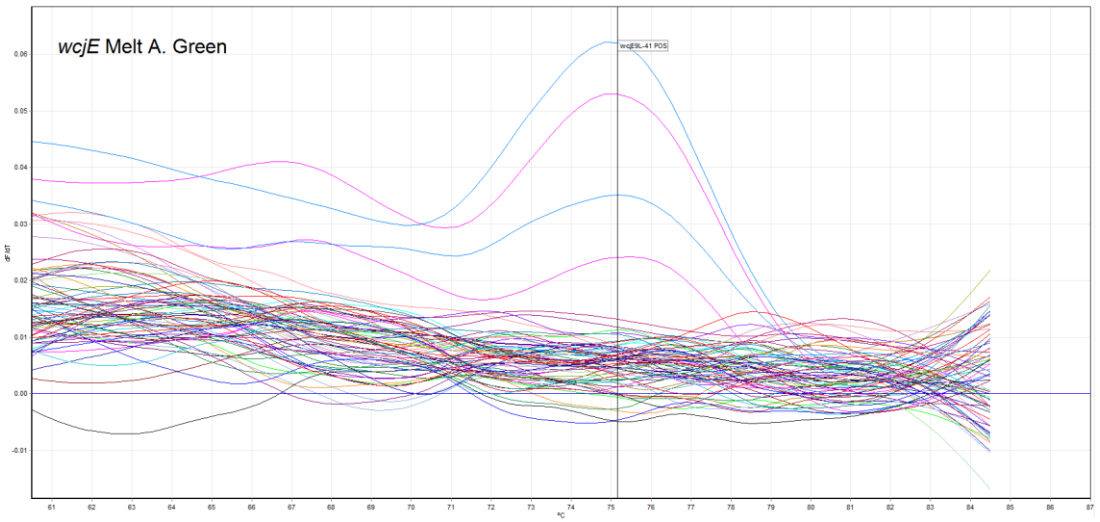

wchA2 HRM Difference
Replicate View
Norm. & minus norm. 1 of wchA2 13 POS
°C

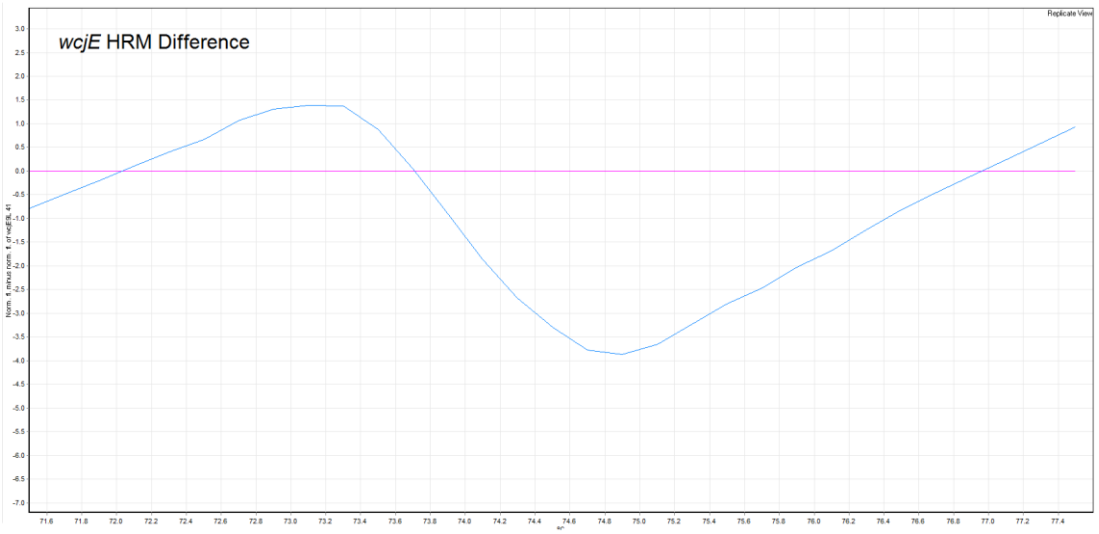

wciK Melt A. Green
wciK 42 POS
-d'F/dT
°C

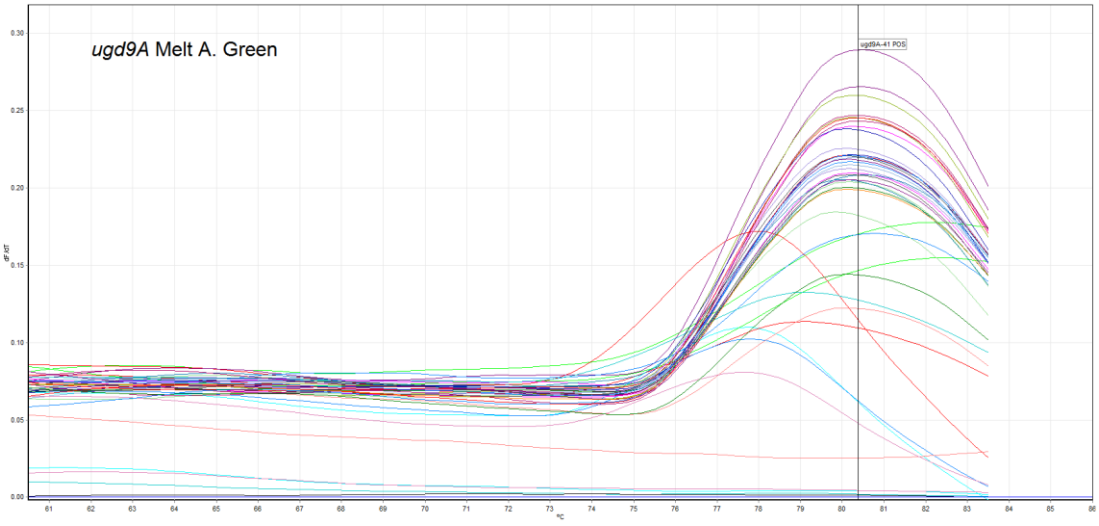

wciK HRM Difference
Replicate View
glf35F Melt A. Green
glf35F HRM Difference
Replicate View

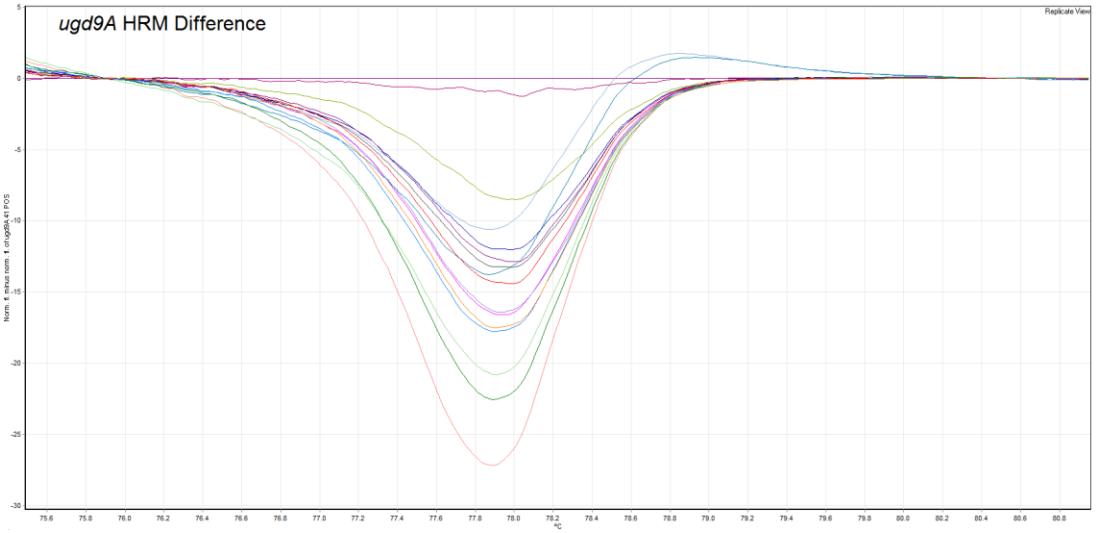

wciN Melt A. Green
wciN6A 24 POS
dF/dT
°C

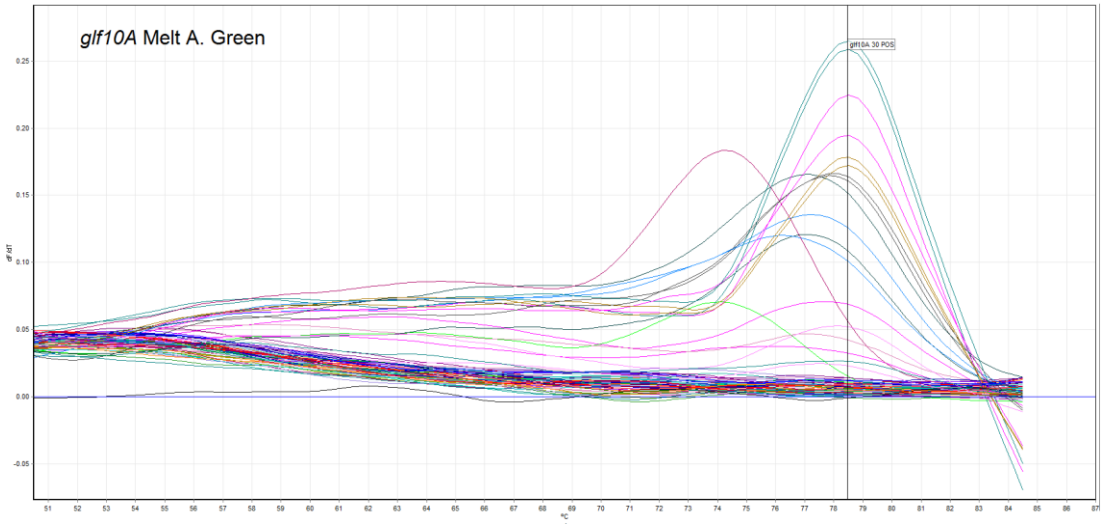

wciN HRM Difference
Replicate View
Norm. fl. minus norm. fl. of wciN6A 24 POS
°C

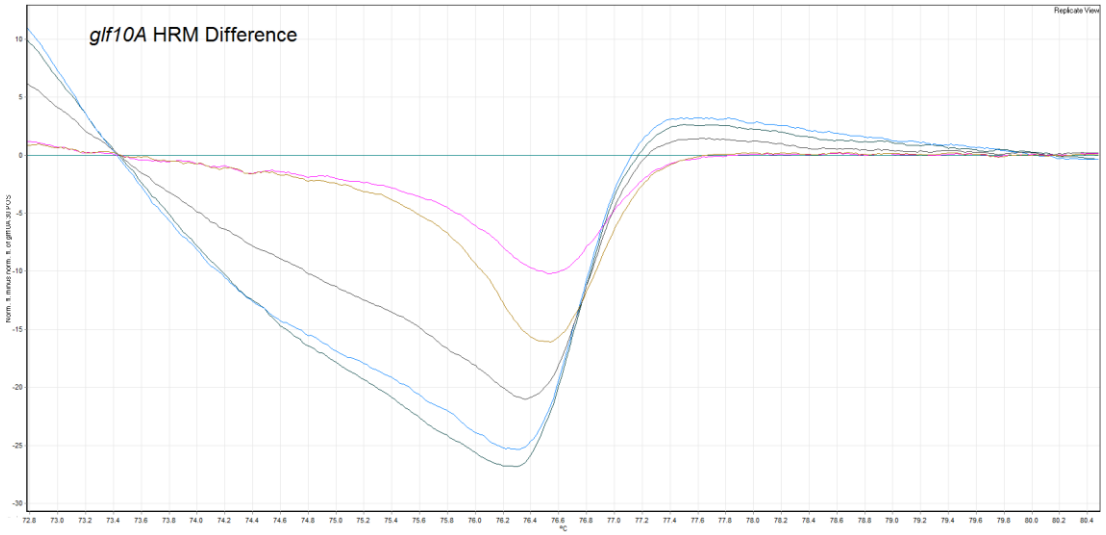

wcjE Melt A. Green
wcjE9L-41 POS
dF/dT
°C

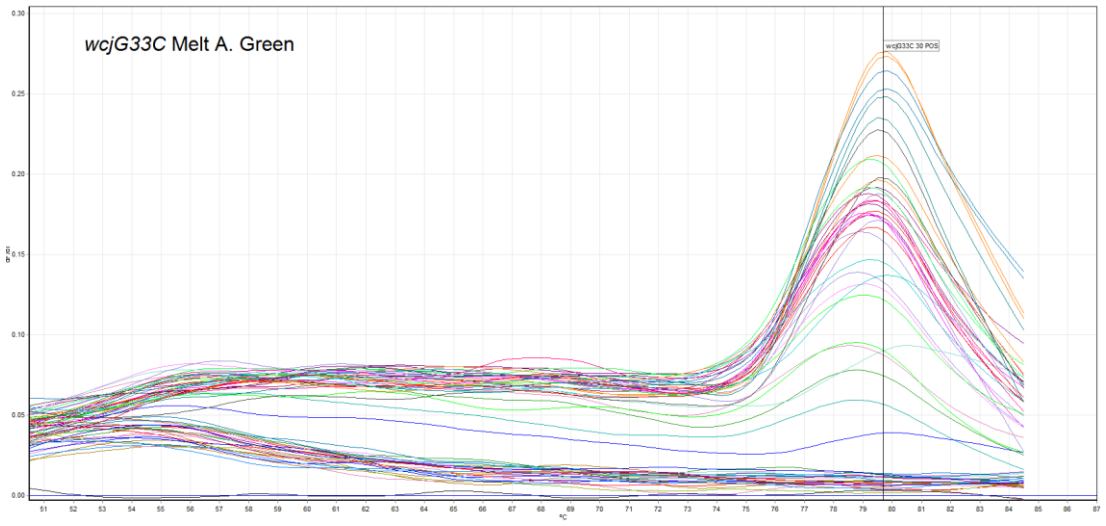

wcjE HRM Difference
Replicate View

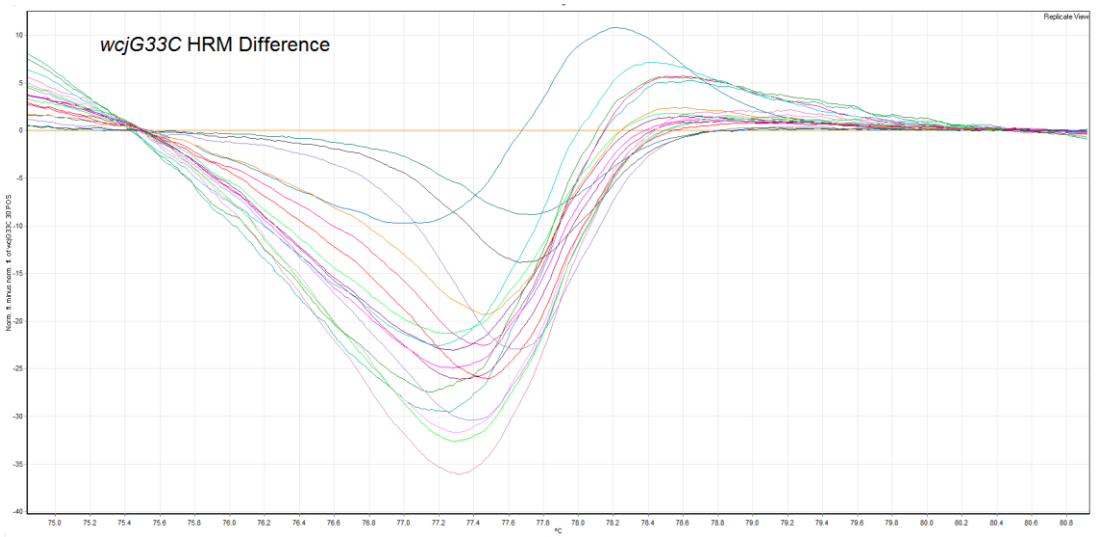

ugd9A Melt A. Green
ugd9A-41 POS

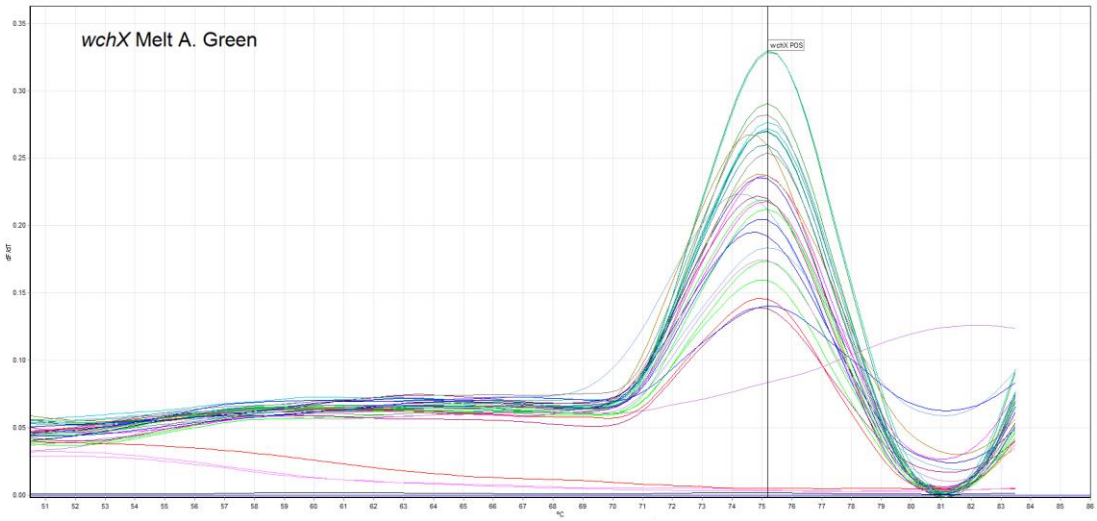

ugd9A HRM Difference
Replicate View

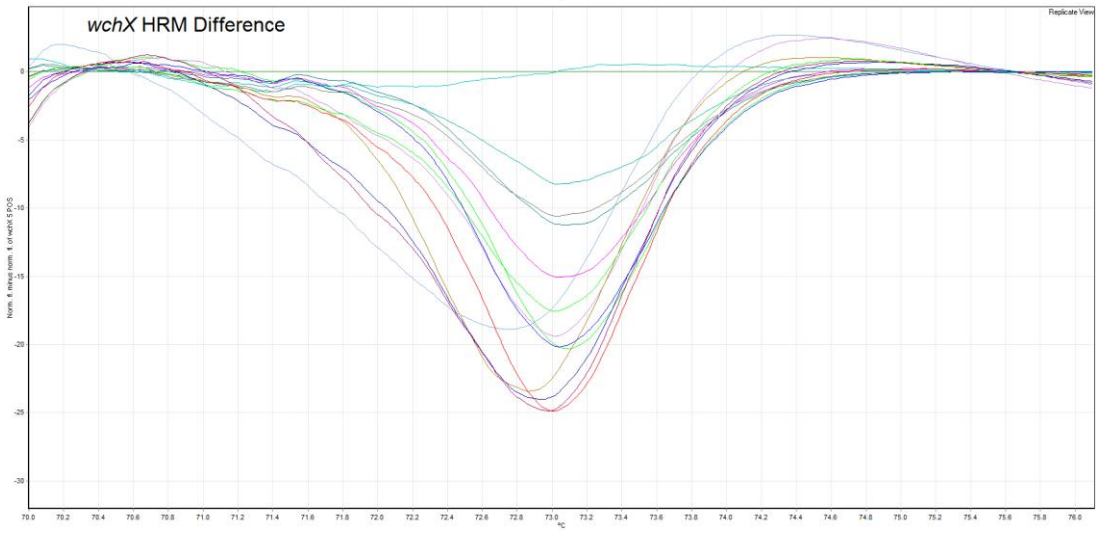

glf10A Melt A. Green
glf10A 30 POS
-d(F)/dT
°C

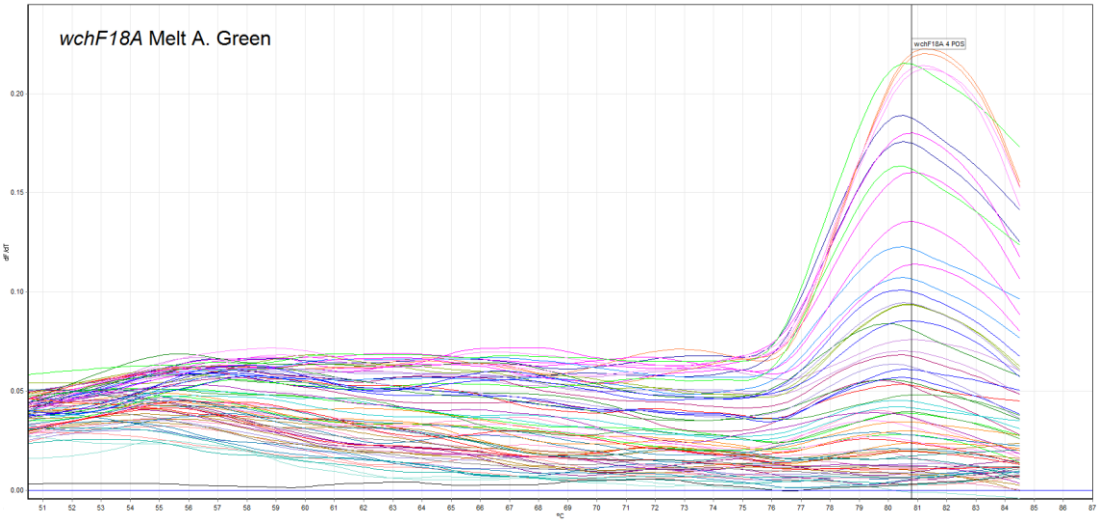

glf10A HRM Difference
Replicate View
°C

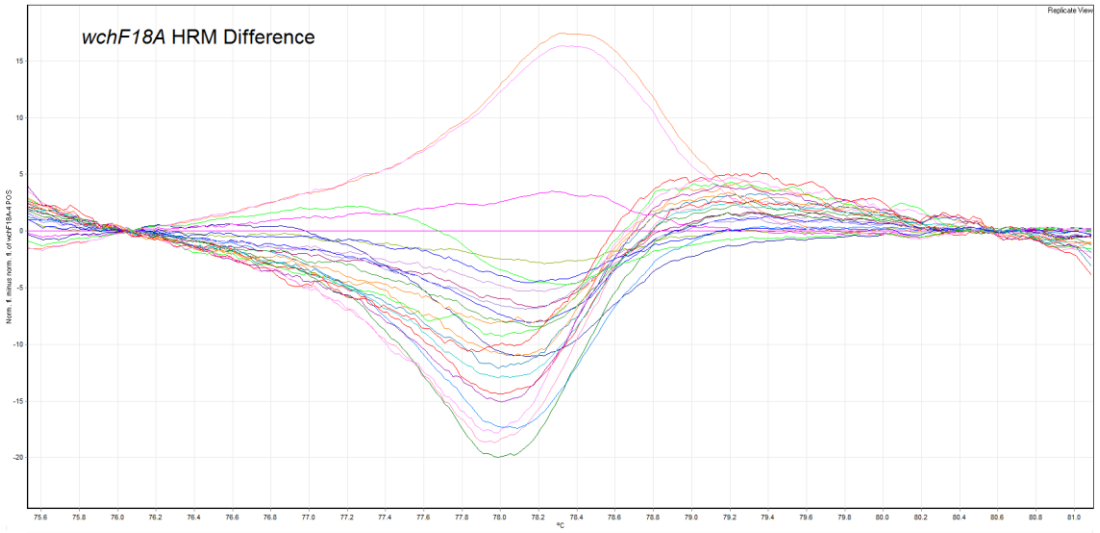

wcjG33C Melt A. Green
wcjG33C 30 POS
-d(F)/dT
°C

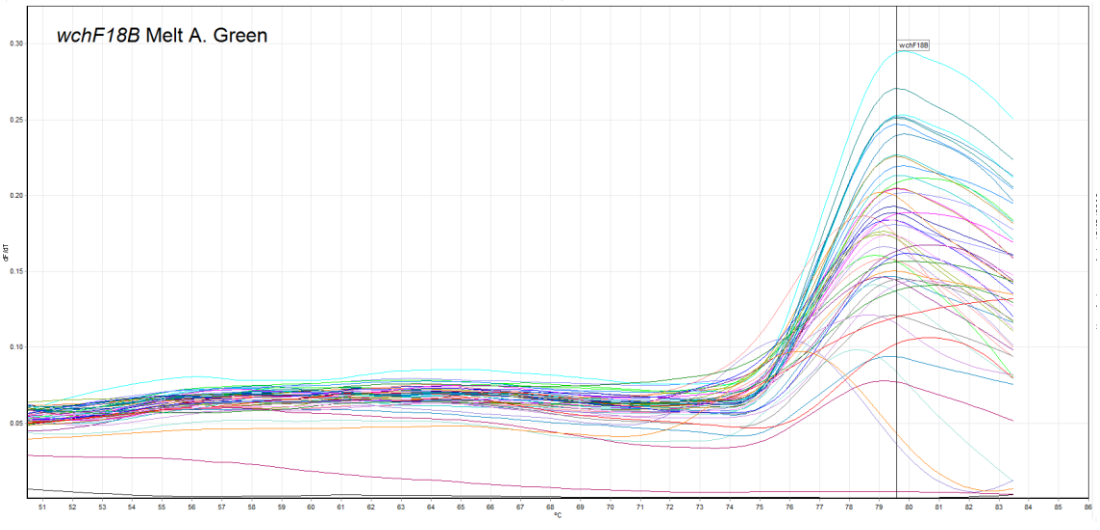

wcjG33C HRM Difference
Replicate View

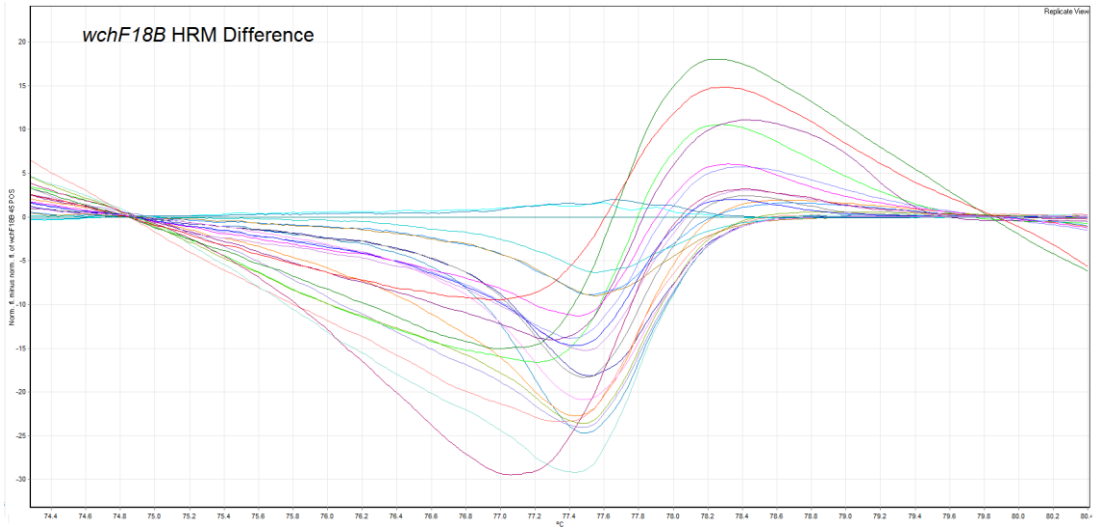

wchX Melt A. Green
wchX POS

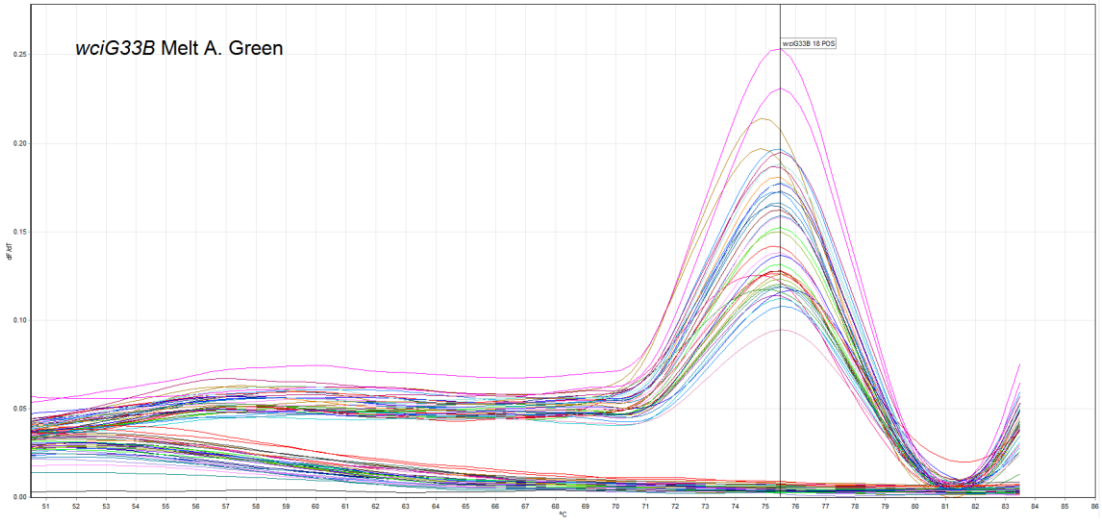

wchX HRM Difference
Replicate View

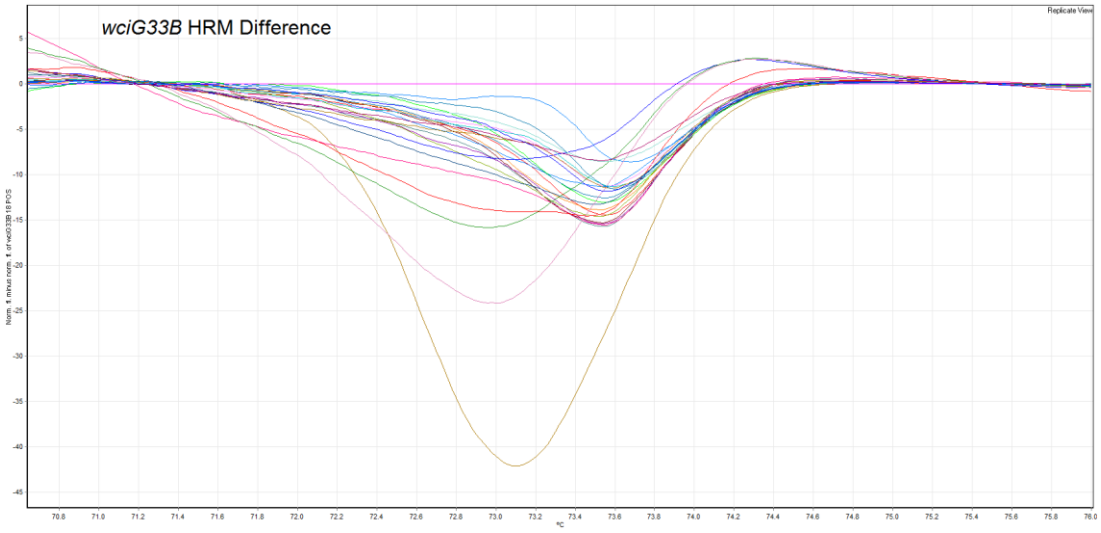

wchF18A Melt A. Green
wchF18A 4 POS
dF /dT
°C


wchF18A HRM Difference
Replicate View
°C


wchF18B Melt A. Green
wchF18B
dF /dT
°C

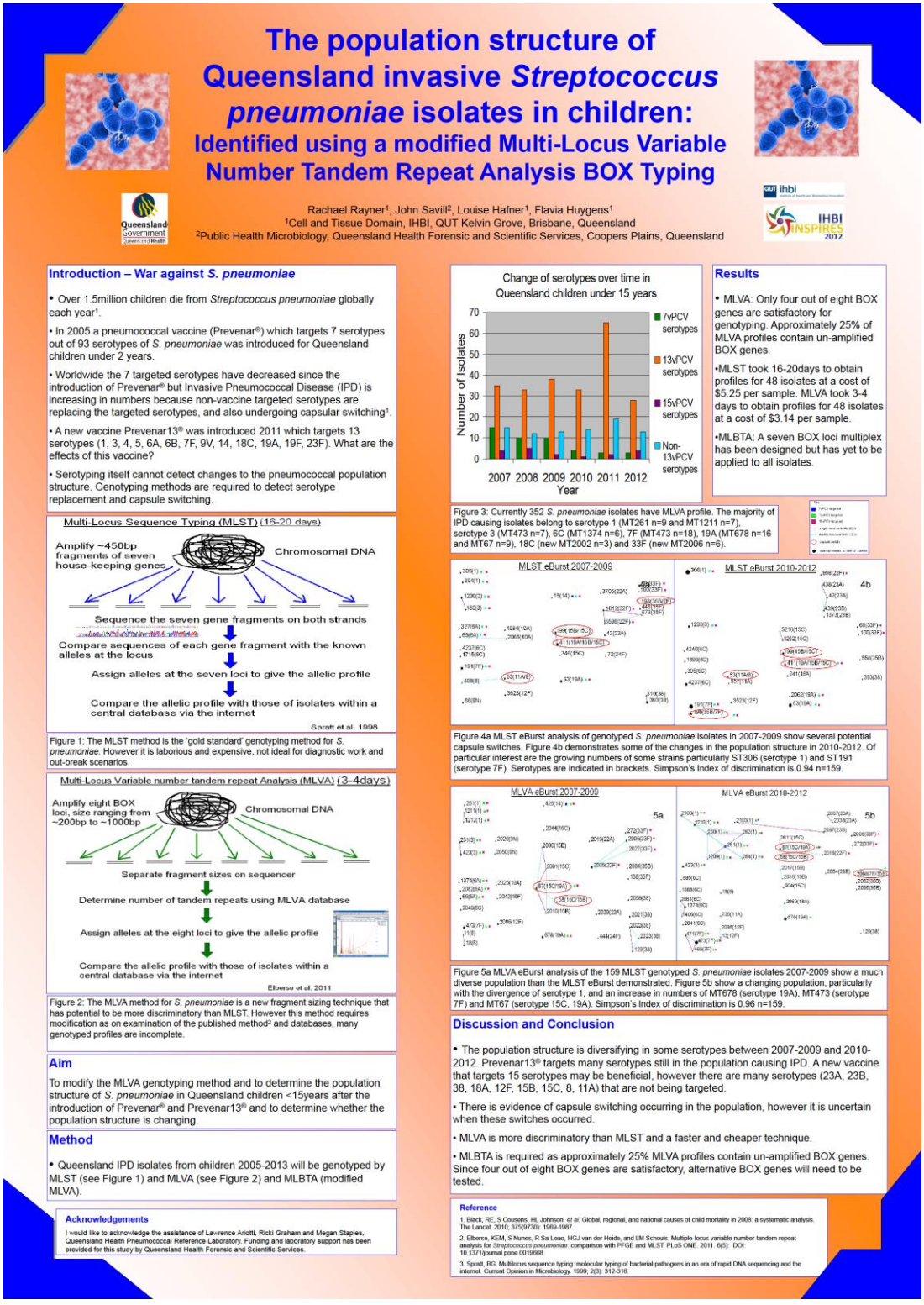

wchF18B HRM Difference
Replicate View


wciG33B Melt A. Green
wciG33B 18 POS

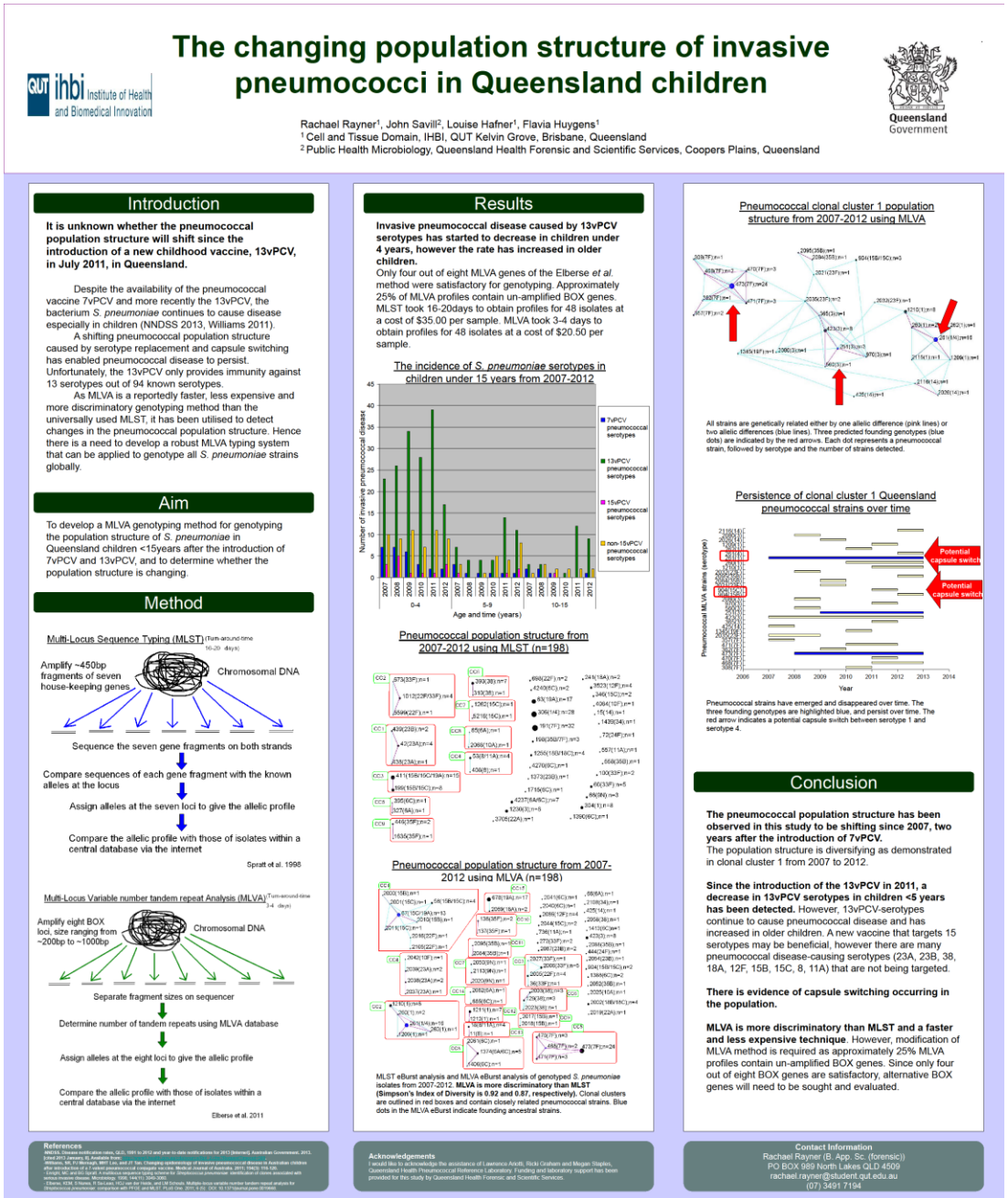

wciG33B HRM Difference
Replicate View

APPENDIX B

Appendix B1: Poster presentation at IHBI Inspires 2012, Gold Coast



Appendix B2: Poster presentation at Australian Society for Microbiology Annual Conference 2013, Adelaide, and Microbiology at QUT 2014, Brisbane



Appendix B3: Oral presentation at IHBI Inspires 2013, Brisbane

QUT ihbi Institute of Health and Biomedical Innovation

ihbi inspires 2013

The War Against *Streptococcus pneumoniae*

Rayner, Rachael[1]; Savill, John[2]; Hafner, Louise[1]; Huygens, Flavia[1]
[1]Cell and Tissue Domain, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD
[2]Public Health Microbiology, Queensland Health Forensic and Scientific Services, Coopers Planes, QLD

Introduction: More than 800 000 children under five globally are dying annually from invasive diseases caused by the gram-positive bacteria *Streptococcus pneumoniae* [1]. Despite the global introduction of a childhood vaccine, the pneumococcus has the ability to evade the effects of the vaccine. Bacterial fingerprinting techniques have been used to characterise the pneumococcus to understand the genetics and relatedness between strains, the sources of infections, the impact of the vaccines and the potential changes that the pneumococcus may undergo. Several fingerprinting techniques are available, however the 'gold standard' is expensive and laborious, whereas other techniques have been developed. A new childhood vaccine, called Prevenar13, was introduced in July 2011 due to increasing pneumococcal diseases. Since the introduction of this vaccine, there have been no reports examining the potential effect the vaccine will have on the pneumococcal population in Queensland children. We aim to develop an inexpensive, highly discriminatory and quick bacterial fingerprinting technique for *S. pneumoniae* and use this to determine the pneumococcal population structure in Queensland.

Methods: A bacterial fingerprinting technique, Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA) was modified and developed from previously published methods [2, 3, 4, 5], and applied to biochemically characterised isolates of invasive *S. pneumoniae* isolated from Queensland children under fifteen years of age from 2007-2012. A published MLVA method [2] and the 'gold standard' Multi-Locus Sequence Typing (MLST) was also applied [6].

Results: A total of 317 pneumococcus isolates were genotyped using our MLVA method. A Simpson's Index of Diversity of 0.98 was achieved and our technique had higher discrimination, was quicker and inexpensive compared to MLST. The majority of pneumococcal isolates belong to serotypes 19A (36%), serotype 7F (9%), serotype 1 (9%), serotype 3 (5%), serotype 19F (4%), serotype 6C (3%) and serotype 33F (3%). Serotype 6C and 33F are not targeted by any current vaccines. Serotype replacement was not observed, however a number of potential capsule switching has been observed.

Conclusions: We have developed our own bacterial fingerprinting method based on MLVA which is more discriminatory, inexpensive and quicker than the 'gold standard' MLST. The Queensland pneumococcal population appears to be diversifying since the introduction of Prevenar13, predominantly due to a reduction in disease.

'Real World' Implications: This study will contribute to the global epidemiology surveillance of *S. pneumoniae*. By developing a highly discriminatory bacterial fingerprinting technique, it can be used globally to enhance our understanding of this common but potentially deadly bacteria. Furthermore, guidance can be given for future childhood vaccines.

References:

[1] Black, RE, S Cousens, HL Johnson, *et al.* Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet, 2010. **375**(9730): 1969-1987.
[2] Elberse, KEM, S Nunes, R Sa-Leao, HGJ van der Heide, LM Schouls. Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. PLoS One, 2011. **6**, e19668 DOI: 10.1371/journal.pone.0019668.
[3] Koeck, J, B Nanjanpop-Lafourcade, S Cade, *et al.* Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. BMC Microbiology, 2005. **5**(1): 66-74.
[4] Van Cuyck, H, B Pichon, P Leroy, *et al.* Multiple-Locus Variable-Number Tandem-Repeat Analysis of *Streptococcus pneumoniae* and comparison with Multiple Loci Sequence Typing. BMC Microbiology, 2012. **12**: 241-258.
[5] Rakov, AV, K Ubukata, D Ashley Robinson. Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus *boxB* sequence typing. Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases, 2011. **11**(8): 1929-1939.
[6] Enright, MC, BG Spratt. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology, 1998. **144**(11): 3049-3060.

Appendix B4: Oral presentation at IHBI Inspires 2014, Gold Coast

QUT ihbi Institute of Health and Biomedical Innovation

IHBIINSPIRES 2014 Postgraduate Student Conference

A novel capsular typing method for *Streptococcus pneumoniae* using Minimum SNPs

Rayner, R E[1]; Savill, J[2]; Hafner, L[1]; Huygens, F[1]
[1] Chronic Disease and Ageing Theme, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland, Australia
[2] Queensland Health Forensic and Scientific Services, Coopers Plains, Queensland, Australia

Introduction: *Streptococcus pneumoniae* is a potentially deadly bacterium that has the ability to switch its protective capsule. Termed capsule switching, this is problematic worldwide since pneumococci can circumvent the current childhood pneumococcal vaccines which targets 13 capsule types out of 95 known capsule types. Determining the numerous capsule switching events that have been observed worldwide in *S. pneumoniae* strains has largely relied on combining serotyping methods with bacterial fingerprinting methods such as Pulsed Field Gel Electrophoresis (PFGE) and Multi-Locus Sequence Typing (MLST). Serotyping the capsule (Quellung or Neufeld) is laborious and expensive, as is whole capsule sequencing, which has been used to verify capsule switches in these strains. Multiplex PCR reactions have been developed for fast and cheap capsule identification but none have been tested on all 95 possible capsule types. This study aims to develop a novel capsule typing method using a bioinformatics approach (Minimum SNPs) to identify polymorphic genes and SNPs within the capsule-cassette. An allele-specific real-time PCR, used in combination with a bacterial fingerprinting method, will be applied to determine capsule switching in pneumococcal strains isolated in Queensland.

Methods: The Minimum SNPs software program was used to analyse 94 different *S. pneumoniae* capsule-cassettes and consequent 'pseudoDNA' sequences to identify a minimum set of targets (capsule genes within the cassette) that characterises each serotype. An allele-specific real-time PCR method was developed based on the selected targets, and applied to clinical isolates of *S. pneumoniae* to determine the capsule-cassette serotype and sub-serotypes in Queensland. Capsule typing was used in combination with Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA) to detect capsule switching within a Queensland population.

Results: *In silico* data analysis identified 18 capsule cassette genes required to distinguish 55 serotypes, and 46 serogroups (Simpson's Index of Discrimination = 0.9936). A further 19 SNPs within capsule genes increased the discrimination to D = 1.0 for all 94 serotypes. Real-time PCR allows fast detection for the presence/absence of a gene or a SNP. In combination with MLVA, pneumococcal capsule switching in Queensland was determined.

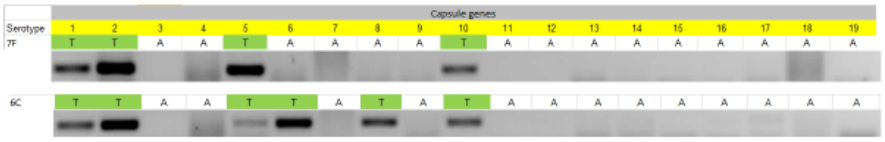


Figure 1: The pseudoDNA sequences for serotype 7F and 6C based on the binary system of T=present and A=absent for each of 19 capsule genes that derives from a 'mother template'. Minimum SNPs program has selected these capsule genes which provide the maximum Simpson's Index of Diversity (D) for 35 serotypes in Queensland.

Conclusions: Identification of capsule switching is important, particularly in light of vaccine evasion by pneumococcal strains. Our study has demonstrated differentiation of the majority of pneumococcal serogroups using a bioinformatics approach. Allele-specific real-time PCR was developed, and combined with MLVA to verify potential pneumococcal capsule switching in Queensland.

'Real-world' Implications: The potential for this inexpensive and quick capsule typing method may enable the rapid detection of pneumococcal capsule switching events worldwide, enabling suitable selection of pneumococcal vaccines.

Appendix B5: Oral presentation at XIX Lancefield International Symposium on Streptococci and Streptococcal Diseases, Buenos Aires, Argentina, and Microbiology in Maleny QLD Branch of ASM Scientific meeting, Maleny, Australia.

A novel capsular typing method for *Streptococcus pneumoniae* using Minimum SNPs

Rayner, RE[1], Savill, J[2], Hafner, L[1], Huygens, F[1]

1 Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland, Australia

2 Queensland Health Forensic and Scientific Services, Coopers Plains, Queensland, Australia

Purpose of Study

Streptococcus pneumoniae is a potentially deadly bacterium that has the ability to switch its protective capsule via the process of horizontal gene recombination. Termed capsule switching, this is problematic worldwide since pneumococci can circumvent the current childhood pneumococcal vaccines. Numerous capsule switching events have been observed worldwide including USA and Italy, and determination of these capsule switches has largely relied on traditional serotyping methods and bacterial fingerprinting methods such as Pulsed Field Gel Electrophoresis (PFGE) and Multi-Locus Sequence Typing (MLST). Serotyping (Quellung or Neufeld) is laborious and expensive, as is whole capsule sequencing, which has been used to verify capsule switches. Multiplex PCR reactions have been developed for capsule identification but none have been tested on all 95 possible capsule types. This study aims to develop a novel capsule typing method using a bioinformatics approach to identify polymorphic genes and SNPs within the capsule cassette. An allele-specific real-time PCR, used in combination with a bacterial fingerprinting method, will be applied to determine capsule switching in Queensland pneumococci.

Method

The Minimum SNPs software program was used to analyse 94 *S. pneumoniae* serotypes and consequent pseudoDNA sequences to identify a minimum set of targets (capsule genes within the cassette) that characterises each serotype. An allele-specific real-time PCR method was developed based on the selected targets, and applied to clinical isolates of *S. pneumoniae* to determine the capsule cassette serotype and sub-serotypes in Queensland. Capsule typing was used in combination with Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA) to detect capsule switching within a Queensland population.

Results

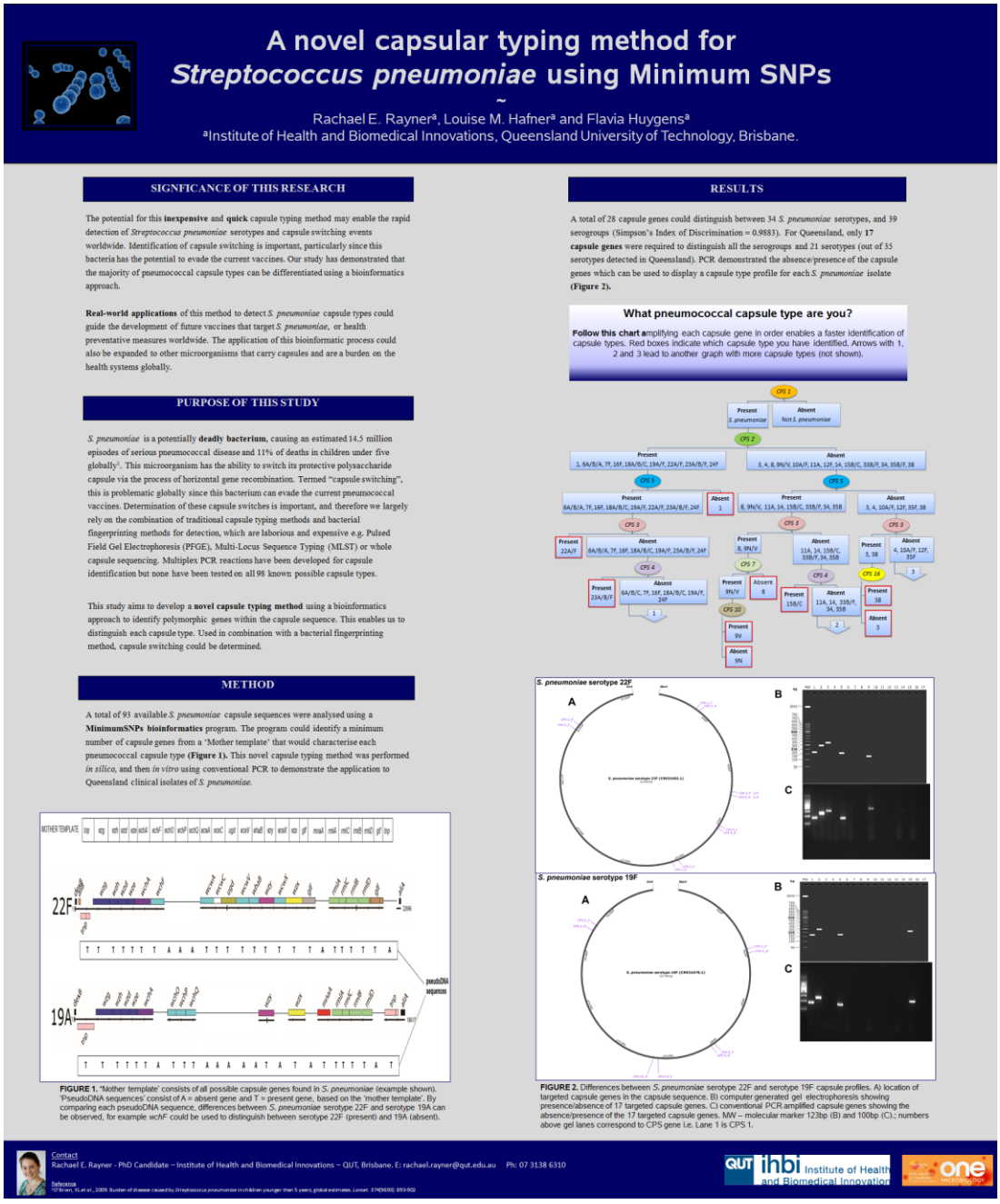
In silico data analysis identified 17 capsule cassette genes required to distinguish 55 serotypes, and 46 serogroups (Simpson's Index of Discrimination = 0.9936). A further four SNPs within capsule genes increased the discrimination to D = 1.0 for all 94 serotypes. Real-time PCR allows fast detection for the presence/absence of a gene or a SNP. In combination with MLVA, pneumococcal capsule switching in Queensland was determined.

Conclusion

Identification of capsule switching is important, particularly in light of vaccine evasion. Our study has demonstrated differentiation of the majority of pneumococcal serogroups using a bioinformatics approach. Allele-specific real-time PCR was developed, and combined with MLVA, potential pneumococcal capsule switching was verified in Queensland. The potential for this inexpensive and quick capsule typing method may enable the rapid detection of pneumococcal capsule switching events worldwide.

Keywords: pneumococcal, capsule, genotyping

Appendix B6: Poster presentation at Australian Society for Microbiology Annual Conference 2015, Canberra



Appendix B7: Statement of Contribution for manuscript "Genotyping *Streptococcus pneumoniae"* (April 2015)

QUT

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Best viewed in Adobe Reader

Statement of Contribution of Co-Authors for Thesis by Published Paper

The following is the format for the required declaration provided at the start of any thesis chapter which includes a co-authored publication.

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter: Chapter 2

Publication title and date of publication or status: Genotyping *Streptococcus pneumoniae*

Published in: Future Microbiology, 2015. Volume 10, issue 4, p. 653-664.

Contributor	Statement of contribution*
Rachael E. Rayner Signature Date 24 July 2015	Wrote the review manuscript, contributed to research.
John Savill 28/7/015	Assisted with writing and drafting the manuscript, and reviewed the manuscript.
Louise M Hafner 29/7/15	Assisted with writing and reviewed the manuscript.
Flavia Huygens	Assisted with writing and drafting the manuscript, and reviewed the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Flavia Huygens
Name

Signature

29/7/15
Date

Appendix B8: Published manuscript "Genotyping *Streptococcus pneumoniae*" to Future Microbiology (April 2015)

REVIEW

For reprint orders, please contact: reprints@futuremedicine.com

Future MICROBIOLOGY

Genotyping *Streptococcus pneumoniae*

Rachael E Rayner[1], John Savill[2], Louise M Hafner[3] & Flavia Huygens*[1]

ABSTRACT *Streptococcus pneumoniae* is a potentially deadly human pathogen associated with high morbidity, mortality and global economic burden. The universally used bacterial genotyping methods are multilocus sequence typing and pulsed field gel electrophoresis. However, another highly discriminatory, rapid and less expensive genotyping technique, multilocus variable number of tandem repeat analysis (MLVA), has been developed. Unfortunately, no universal MLVA protocol exists, and some MLVA protocols do not amplify certain loci for all pneumococcal serotypes, leaving genotyping profiles incomplete. A number of other genotyping or characterization methods have been developed and will be discussed. This review examines the various protocols for genotyping *S. pneumoniae* and highlights the current direction technology and research is heading to understand this bacterium.

KEYWORDS
• epidemiology
• genotyping • MLST • MLVA
• *Streptococcus pneumoniae*

Streptococcus pneumoniae is a potentially deadly human pathogen associated with high morbidity (estimated 14.5 million episodes of serious pneumococcal disease in the year 2000), high mortality (causes 11% of deaths in children aged under 5 years) and high economic burden globally, especially in underdeveloped countries [1]. Since the introduction of a childhood pneumococcal vaccine (7-valent Pneumococcal Conjugate Vaccine [7vPCV], Wyeth, NJ, USA), serotype replacement of targeted pneumococcal serotypes with nontargeted pneumococcal serotypes have almost countermanded the effect of the vaccine [2,3]. As well as this, *S. pneumoniae* has the ability to switch their protective polysaccharide capsule with other pneumococci, enabling both to carry a nonvaccine targeted capsule and become immune to the vaccines, leading to 'vaccine-escape' strains [4–6]. As of December 2012, a total of 86 (44%) WHO member states have added a PCV to their routine infant immunization schedule of the national immunization programs [7].

Characterization of bacteria below the species level using subtyping methods enable the assessment of the impact of human interventions such as vaccines and antibiotics, the relatedness of bacterial isolates and the sources and transmission routes of infections [8]. Traditional characterization subtyping techniques, such as Quellung serotyping, has been used for decades in pneumococcal epidemiology studies [9]. Unfortunately, only 95 published serotypes can be determined (compared with thousands of genotypes identified within these 95 serotypes), not all serotypes can be distinguished (serotype 6A and 6C are distinguished through PCR) and capsule switching cannot be detected without the assistance of a genotyping method. The suitability of genotyping tools for epidemiology studies depends on its ability to distinguish different fingerprint patterns for unrelated isolates from

[1]Institute of Health & Biomedical Innovation (IHBI), 60 Musk Ave, Kelvin Grove, 4059, Queensland, Australia
[2]Public Health Microbiology Laboratory, Queensland Health Forensic & Scientific Services, Coopers Plains, Queensland, Australia
[3]Queensland University of Technology (QUT), Brisbane, Queensland, Australia
*Author for correspondence: Tel.: +617 3138 0453; f.huygens@qut.edu.au

Future Medicine part of fsg

10.2217/FMB.14.153 © 2015 Future Medicine Ltd *Future Microbiol.* (2015) 10(4), 653–664 ISSN 1746-0913

identical patterns for related isolates [10]. As there are 95 serotypes of *S. pneumoniae* and thousands of different genotypes, highly informative tools for analysis are required [11–16].

The current 'gold standard' genotyping method for *S. pneumoniae* is multilocus sequence typing (MLST) [17], superseding pulsed field gel electrophoresis (PFGE) [18,19]. Both methods have high discriminatory power, good resolution and good reproducibility; however limitations still remain. Multilocus variable number of tandem repeat analysis (MLVA) is another promising genotyping method. Unfortunately, there are many different MLVA protocols for *S. pneumoniae*, and limitations have been identified [8,20–25]. Other characterization methods have been developed, including serotyping [9], mass spectrometry (MS) [26,27] and next-generation sequencing (NGS) [28,29]. This review outlines the current *S. pneumoniae* genotyping methods and highlights the importance of further development of a robust and universal genotyping method.

Traditional characterization methods

Traditional characterization methods for epidemiology studies, such as serotyping (Quellung or Neufeld), only provides a broad picture of changes, cannot differentiate genotypes or capsule switching, and is restricted to large reference laboratories due to costs and availability of specific reagents [30]. In the Quellung reaction, antibodies bind and react to the pneumococcal polysaccharide capsule, causing it to become opaque and enlarged when visualized under a microscope [9]. However, phenotypic variations of the polysaccharide capsule do not necessarily reflect genetic variations because external influences can change the phenotype [10]. Therefore a number of PCR-based serotyping methods have been developed for *S. pneumoniae* [31–44]. Despite numerous PCR-based serotyping methods, the Quellung reaction has remained as the 'gold standard' characterization technique. Unfortunately, it is also commonly the sole subtyping method used in pneumococcal epidemiology studies, thus posing limitations on the comprehension of the relatedness between two invasive isolates at the genetic level, identifying specific sources of infections (apart from a general outbreak in a localized area) and the transmission of specific pneumococci across states and countries, and fails to identify capsule switching without using a genotyping method.

MLST & PFGE: current 'gold standard' genotyping methods

MLST is considered the 'gold standard' genotyping method for characterizing *S. pneumoniae*. It is used by the Pneumococcal Molecular Epidemiology Network (PMEN), an association which aims to survey the global antibiotic-resistant *S. pneumoniae* and standardize nomenclature and classification of resistant clones (web1. sph.emory.edu/PMEN/) [45]. Developed by Enright and Spratt [17], there are over 9800 different MLST genotypes identified from over 24,000 entries submitted to the international online database [45,46].

MLST has the ability to produce unambiguous results, is portable, provides good resolving and discriminatory power that can be used for local and global epidemiology, and can be directly applied to samples (e.g., cerebral spinal fluid) without the need for culture [17,23,47–49]. Derived from multilocus enzyme electrophoresis [48], MLST utilizes seven housekeeping genes (*aroE, gdh, gki, recP, spi, xpt* and *ddl*) to genotype *S. pneumoniae* [17]. Housekeeping genes are considered stable and less prone to recombination than other regions of the *S. pneumoniae* genome [17]. These conserved regions are considered more appropriate for evolutionary rather than epidemiology studies where frequently changing genomic regions, such as those used in MLVA, are more desirable [30]. A fluorescent label is incorporated into PCR-amplified housekeeping genes for sequencing. Single nucleotide polymorphisms (SNPs) distinguish alleles and a sequence type (ST) is generated from the combined seven alleles. Unfortunately, MLST also poses limitations including the high expense to genotype, especially when examining large batches of isolates, it is technically demanding with several PCR steps, DNA purification steps and sequencing analysis, and it has not been developed for routine use despite automation [20,23,50–52]. Data analysis can also be laborious and can take longer than the analysis of MLVA data [30].

On the other hand, PFGE has higher discriminatory power than MLST, good resolution and good typeability [17,23,47–48]. PFGE was faster and more discriminatory than multilocus enzyme electrophoresis, prompting its wide use before MLST was developed [53]. Restriction enzymes are used to digest genomic DNA of *S. pneumoniae* into large fragments which are separated by pulsed gel electrophoresis; the

DNA banding patterns (fingerprints) are then compared between isolates [18,51]. PFGE is listed as a standard genotyping method by PMEN. However, it would no longer be considered a 'gold standard' genotyping method for *S. pneumoniae* considering the small size and quality of the database in comparison to the international MLST database, the lack of a universal and generally accepted scheme, the fact that PFGE suffers from lack of portability between laboratories unless standardized, is ambiguous, is not amenable for international database, is laborious and poses a potential health hazard through prolonged handling of cultivable strains [23,50,54–55]. As a result, MLST became a more popular genotyping method. Even MLVA can be more discriminatory than PFGE, where four to five highly variable genomic regions have been used to characterize *Salmonella enterica* serovar Typhimurium infections [24].

Despite the popularity of MLST, researchers have still been testing for improvements, particularly to minimize technical requirements, cost and laboratory work. Methods that are relatively cheap and applicable in standard procedures on mass scale are more desired than expensive technology with high differentiation [30]. Crisafulli *et al.* [56] developed an alternative MLST method which utilizes 96 loci (rather than seven) based on 39 complete *S. pneumoniae* genomes. By increasing the number of loci, strains with the same ST but different capsule types can be differentiated, and a higher discrimination and better resolution can be achieved. However, sequencing 96 loci is quite laborious, as their method indicates that each isolate requires a full 96-well plate to obtain a complete profile; consequently, only 69 isolates were sequenced. This paper claims that the 96-MLST could still be applicable as a high-throughput method, even in small-scale laboratories, with costs per isolate comparable with NGS methods. It has since been applied by Moschioni *et al.* [57] to characterize a large MLST clonal cluster 156.

MLST & NGS

With the advancement of NGS technology, a number of protocols combining MLST with NGS have been published. MLST has been modified by Boers *et al.* [28] who incorporated NGS technology such as the Roche 454 sequencing to provide a high-throughput MLST (HiMLST) method. HiMLST attaches a unique DNA barcode called the multiplex identifier (MID), allowing the combination of multiple species of bacteria. They have simultaneously sequenced *Legionella pneumophila, Staphylococcus aureus, Pseudomonas aeruginosa* and *S. pneumoniae* [28]. HiMLST provides a high-throughput and cost-effective modification of MLST (tenfold reduction in costs using HiMLST method compared with Sanger sequencing of MLST genes) especially if using NGS technology. However in comparison to current MLVA techniques, HiMLST is still expensive and time consuming. HiMLST could be adapted to become more discriminatory than MLST by increasing the number of targets [28]. A similar NGS technique called MLST-seq uses species-specific hairpin primers and MIDs to amplify 20 target loci in *Salmonella enterica* with Roche 454 sequencing. Nevertheless, a number of targets were missed due to uneven sequence coverage, and it has not been applied to *S. pneumoniae* although the principles should be transferable [28,29].

MLST & MS

Other MLST genotyping techniques have been devised utilizing MS, which has been reported to be the 'next generation tool' for identifying species of various microorganisms [58]. Matrix-assisted laser desorption ionization time-of-flight mass-spectrometry (MALDI-TOF-MS) has been used to identify *S. pneumoniae* species, as well as characterize isolates further using MLST housekeeping genes, although there have been reports that identification can be difficult [26,59–60]. Papua New Guinean *S. pneumoniae* isolates were genotyped using cleaved MLST housekeeping genes that had mass signal patterns analyzed using MALDI-TOF-MS [26]. This MS-based method requires additional sample preparation; however the speed of analysis allows this method to be more cost-effective than MLST. MALDI-TOF-MS reported 99.0% concordance rate with MLST is faster and more cost effective than MLST sequencing but manual inspection was required for 40% of total alleles.

An alternative genotyping method, pneumococcal serotyping and genotyping method (PSGS), amplifies 40 genes (32 serotype-specific genes located within the capsular biosynthetic loci, and eight MLST housekeeping genes) using PCR and electrospray ionisation mass spectrometry (PCR/ESI-MS) [27]. PSGS is reported to have less resolving power than MLST [27].

MS technology could be used for analysis of MLVA loci in future studies, however PCR fragment sizes of 100–200 bp should be used because of increased difficulty to produce single-stranded products in gas phase for larger fragment sizes [61]. MLVA products for *S. pneumoniae* are generally greater than 200 bp, and reach up to greater than 1000 bp [8]. The feasibility of combining MLVA with MS will have to be investigated.

MLVA: the emerging genotyping method

It is the general consensus that the most likely subtyping method will be based on genomic sequencing, which MLST currently provides when characterizing *S. pneumoniae* strains. However even though the expensive cost and speed of sequencing larger fragments of DNA is continuously decreasing, there is still the cumbersome burden of processing and analyzing these data into meaningful information. Therefore it is sometimes necessary to only examine the genomic regions that are the most useful, which may mean only examining certain genes rather than the entire genome.

As a consequence, MLVA was first developed by Koeck *et al.* [20] in 2005 as an alternative genotyping method for *S. pneumoniae*. MLVA is claimed to be more discriminatory than MLST, therefore it may be more suitable for local outbreaks of diseases [8,22,62]. However, since MLVA genotyping is not known to be used by the PMEN and has only been used by a few research groups globally, MLST is currently used as a comparative genotyping method.

MLVA amplifies variable number of tandem repeats (VNTRs) instead of housekeeping genes to form an allelic profile [8]. VNTRs are preferred for epidemiology studies and outbreaks due to the high variability in the genes. However this can also lead to instability of some loci, therefore it has been recommended that further studies are required to thoroughly understand these mutations and mutation rates [30]. Multiplexing with fluorescent labels can be utilized and PCR products are separated by gel or capillary electrophoresis according to fragment size which is assigned to specific allele type (**Table 1**) [8]. Using capillary electrophoresis enables automation and a higher throughput with the use of an internal size ladder for accurate sizing [8]. It is unknown whether VNTRs are under neutral or positive selection, however environmental influences can affect the speed of VNTR changes [63]. Recombination rates in pneumococci are high [64]. VNTRs are polymorphic, making them "suitable target(s) for assessing genetic polymorphism(s) within bacterial species" [20,55]. MLVA has been applied to many bacteria including *Bacillus anthracis, Staphylococcus aureus, Yersinia pestis, Salmonella typhi* and *Escherichia coli* O157 [23,51,55].

The original MLVA protocol for *S. pneumoniae* included amplifying 17 VNTRs [20]. This method has been used to genotype 48 isolates in Burkina Faso from 2002 to 2005 [62]. Pichon *et al.* [21] modified this MLVA protocol to distinguish the relatedness of an outbreak of serotype 5 in a northeast London suburb in the winter of 2007–2008 by selecting VNTR loci with the highest diversity based on Simpson's Index of Diversity (namely ms17, ms19, ms25, ms34, ms36, ms37 and ms39).

Another MLVA technique was developed by Elberse *et al.* [8] which used eight BOX loci selected from an original 13 randomly selected BOX loci from the *S. pneumoniae* R6 genome. BOX elements are VNTRs found within intergenic regions of the *S. pneumoniae* genome, comprising three segments: *boxA* (59 nucleotides), *boxB* (45 nucleotides) and *boxC* (50 nucleotides), with *boxB* containing tandem repeats [50,66]. *BoxA* and *boxC* are highly conserved between multiple species of bacteria, however. *boxB* is only observed in *S. pneumoniae* [9,66–68]. BOX loci have varying numbers of repeat regions, for example, BOX-04 has the highest reported variation in repeat numbers (0–17) while BOX-11 has the lowest variation (1–2), allowing for high polymorphism and diversity [8]. BOX loci remain stable under laboratory conditions, can form stable stem-loop structures and most are located near virulence genes (*neuA* and *ply*) or transformative genes, indicating they could be regulatory elements [8,21,66].

Elberse's MLVA include BOX-01 (Spneu40), BOX-02 (Spneu32), BOX-03 (Spneu15), BOX-04a, BOX-04b (Spneu33), BOX-06a, BOX-06b (Spneu38), BOX-11, BOX-12 (Spneu37) and BOX-13 (Spneu25), and are amplified in two multiplex PCRs with fluorescently labeled probes (FAM, NED, VIC and PET) and sized on an automated DNA sequencer [8,37,69]. This method has been standardized so that different platforms produce comparable results [8,30]. MLVA standardization for *Shigella sonnei* utilized calibration strains and has been implemented to participating laboratories for data normalization [25,70]. Elberse's MLVA has been applied to 1154 isolates of invasive pneumococci in the Netherlands, which correlated to 444 MLVA types of *S.*

Table 1. Assignment of allele number to multilocus variable number of tandem repeat analysis fragment sizes for BOX-02 and BOX-03[1].

BOX locus	Allele number	Fragment size (bp)	Left binning	Right binning	Comments
BOX-02	0	279.0	3	3	
	1_01	286.0	3	3	Deletion in *boxA*, allele contains one repeat
	1	322.5	3	3	
	2_01	330.0	3	3	Deletion in *boxA*, allele contains two repeats
	2_02	359.0	3	3	Deletion in *boxC*, allele contains two repeats
	2	366.0	3	3	
	3	411.0	3	3	
	4	454.5	4	4	
	5	499.0	4	4	
	6	545.2	4	4	
BOX-03	1	393.0	3	3	
	2	435.4	3	3	
	3	480.0	3	3	
	4_01	490.0	3	3	Deletion in *boxA* and part of *boxB*, allele contains four repeats
	4	525.0	3	3	
	5	570.0	3	3	
	5_01	578.0	3	3	Insertion in *boxA*, allele contains five repeats
	6	614.3	4	4	
	7	657.0	4	4	
	8	702.0	4	4	
	9	747.0	5	2.5	
	9_01	753.0	2.5	5	To be sequenced
	10	789.0	5	5	
	11	834.0	3	5	
	12_01	860.7	5	5	To be sequenced
	12	879.0	5	5	
	13	918.0	5	5	
	14	963.0	5	5	
	15	1003.2	5.5	5.5	
	16	1048.8	5	5	

[1]Data taken from [8].
Obtained from the MLVA database [65].

pneumoniae. Pichon *et al.* [21] was able to determine nine distinct MLVA types of serotype 5, which were associated with three MLST types, highlighting that MLVA has higher discriminator power. Elberse *et al.* [8] has claimed that their MLVA with eight loci provides high resolution and have high congruence with MLST, and is more time efficient and less laborious than Koeck's 17 singleplex MLVA **(Table 2)**.

Unfortunately, Elberse's results, published on the MLVA database [65] contained many typing failures which were assigned '99' where no PCR amplification was observed [74]. Examination of the MLVA database reveals that 36.5% MLVA profiles (530/1450) contain at least one unamplified locus, and at least 10% of profiles (146/1450) contain two or more unamplified loci. Similarly, Koeck's MLVA database [72] also contains '99' in some genotypes [75]. Primers may have been designed for universal use, and are failing to amplify specific serotypes, for example, BOX-06 in serotype 7F due to the possibility that the locus is missing [69] (unpublished data). Elberse *et al.* [8] have already described a failure to amplify BOX-06 in 89% of serotype 7F. As such, they have still assigned an MLVA type (MT) to these strains with a failed BOX-06. It is unknown whether these serotype 7F strains are all genetically identical or whether some of these strains may be single locus variants (SLV) within BOX-06. This issue remains unresolved, although sequencing of unusual size fragments has been reported [76]. As there are 127 BOX genes in *S. pneumoniae* TIGR4 strain and 115

Table 2. Comparison of multilocus variable number of tandem repeat analysis, multilocus sequence typing, multi-locus *boxB* typing and pulsed field gel electrophoresis for genotyping *Streptococcus pneumoniae*.

Components of genotyping method	MLST protocol	MLVA protocols				PFGE protocol	Ref.
	MLST [17]	*Elberse* et al. [8]	*Koeck* et al. [20]	*Van Cuyck* et al. [22]	*MLBT Rakov* et al. [23]	*PFGE*	
DNA	Seven housekeeping genes	Eight BOX loci	Sixteen VNTR	Seven VNTR	Ten BOX loci	Whole genome	[8,17,20,22–23]
Simpson's index of diversity (S)	0.987 (263 isolates)	0.993 (263 isolates)	0.995 (52 isolates)	0.989 (331 isolates)	0.973 (203 12F isolates)	0.985 (263 isolates)	[8,20,22–23]
Unambiguous	Yes	Yes	Yes	Yes	Yes	No	[8,17,45,51]
Automation	Yes	Yes	Yes	Yes	Yes	No	[55]
International database (sighted 20 February 2014)	Yes (22,972 profiles) [71]	Yes (1786 profiles) [65]	Yes (1087 isolates) [72]	No	No	No	[73]
Portable	Yes	Yes	Yes	Yes	Yes	Limited	[8,17,47]
Resolution	High	Unknown	Unknown	Unknown	Unknown	Good	[17,47]
Sensitivity	High	High	High	High	High	High	[49]
Reproducibility	Good	Good	Unknown	Unknown	Unknown	Unknown	[8]
Technical demands	High	Minimal	Unknown	Unknown	Unknown	High	[8,51]
Time	Long	Quick	Unknown	Unknown	Unknown	Long; laborious	[8,51,55]
Cost	Expensive	Inexpensive	Unknown	Unknown	Unknown	Inexpensive	[8,51]

MLBT: Multi-locus boxB typing; MLST: Multilocus sequence typing; MLVA: Multilocus variable number of tandem repeat analysis; PFGE: Pulsed field gel electrophoresis.

BOX genes in *S. pneumoniae* R6 strain [64], it is important that the selection of BOX loci to genotype *S. pneumoniae* is universal between all isolates, are highly discriminatory and are easily amplified using PCR. A number of microbial typing software have been published, allowing the selection of optimum targets for microbial strain typing [7,77–78].

In the meantime, to increase discrimination of MLVA methods, Rakov *et al.* [23] devised multi-locus *boxB* typing (MLBT) to genotype *S. pneumoniae* serotype 12F clonal complex (CC) 218 by amplifying ten BOX loci (seven used by Koeck *et al.* [20] and three used by Elberse *et al.* [8]). One additional BOX locus was used by Rakov *et al.* [23], namely B10. MLBT combines the allelic size differences of each BOX fragment as well as examining SNP differences in each tandem repeat, increasing the discriminatory power. MLBT was reported to have high discrimination compared with MLST, however no comments were made comparing MLBT with MLVA. From observing the results, MLBT produced a vastly complicated set of results that would not contribute to the established MLVA or MLST databases. However, if the future of subtyping is in genome sequencing then MLBT demonstrates that the BOX genes can be utilized.

More recently, an updated MLVA method of Koeck *et al.* [20] has been published by Van Cuyck *et al.* [22] which uses seven VNTR loci (namely ms17, ms19, ms25, ms33, ms37, ms39 and ms40) based on a selection of 14 VNTR loci. These seven MLVA targets were selected on the basis of a Hunter-Gaston Diversity Index of >0.8 [22]. This updated method has been applied to 331 invasive isolates from 2002 to 2006 in the United Kingdom before the introduction of a childhood pneumococcal conjugate vaccine (PCV). From these data, only ten MLST types have been genotyped – namely ST65, ST138, ST156, ST162, ST176, ST180, ST199 and ST306. A recent study in Australia (submitted paper) has shown that this method failed to amplify ms19 in serotype 3 (ST1230) and serotype 38 (ST393 and ST310), which has not been detected in the UK study. It is known that pneumococcal populations differ between countries; therefore the selection of these seven MLVA markers may not be suitable for other populations. Van Cuyck *et al.* [22] have failed to comment on the nonamplified loci observed by Elberse *et al.* [8], particularly since three of the VNTRs selected are the same as in the Netherlands study.

Van Cuyck *et al.* [22] have also claimed that the ST and serotype can be deduced from the MLVA type, depending on the clonality of the

serotypes without actually having to perform MLST or serotyping. This assumption may not be applicable for all serotypes, as they have also demonstrated that not all ST of serotype 19F could be inferred from the MLVA type despite a high congruence between MLST and MLVA methods [22].

Due to the ability to amplify and a high level of discrimination, a number of loci have consistently been used in the five different modifications of MLVA. Spneu25 (BOX-13) has been adopted in all the MLVA techniques, particularly since it has one of the highest discriminatory power **(Table 3)**. Spneu17 is recorded with the highest discrimination (Hunter-Gaston Diversity [DI]: 0.883) in Koeck *et al.* [20] while it was a close second behind Spneu37 (BOX-12) in Van Cuyck *et al.* [22] study. Elberse *et al.* [69] also described a DI>0.8 for BOX-12 however their highest discriminatory locus was BOX-04 (Spneu33). The locus with the lowest discrimination varied between methods with BOX-11 (DI: 0.460), Spneu41 (DI: 0.484) or Spneu38/BOX-06 (DI: 0.557). It appears that the discriminatory power of each locus varies slightly in different pneumococcal populations. Overall, the following eight loci have obtained a DI >0.8 in at least one MLVA technique: Spneu15 (BOX-03), Spneu17, Spneu25 (BOX-13), Spneu33 (BOX-04), Spneu36, Spneu37 (BOX-12), Spneu39 and Spneu40 (BOX-01) **(Table 3)**. This indicates that a combination of these loci may provide a highly discriminatory MLVA method that could be applied universally to genotype *S. pneumoniae*. Increasing resolution by using more VNTRs has to be weighed against costs [30].

Feasibility of applying MLVA

Advantages of genotyping *S. pneumoniae* with MLVA include that it is more discriminatory, simple, fast to perform, less expensive than MLST and is easily applicable to outbreak investigations [8,23,55]. Other advantages include minimal technical expertise required, the ability for automatic analysis by computer software, unambiguous results and provides robust intralaboratory comparison [23,51]. There is high congruence between MLVA and MLST, indicating that MLVA would be a suitable and reliable technique that could be used instead of MLST [8,22].

The main disadvantage of MLVA currently is the fact that there is no universally accepted technique. This has limited the use of MLVA genotyping applied to study pneumococcal populations and outbreaks. Comparison of current MLVA *S. pneumoniae* genotypes between countries is not possible since each MLVA protocol has assigned a different number of loci to distinguish each strain.

Other *S. pneumoniae* characterization techniques

Other developed characterization techniques have assisted in the epidemiology study of *S. pneumoniae*. BOX-fingerprinting uses BRA and BRB primers to generate a 151bp BOX repeat gene fragment in the *S. pneumoniae* genome [68]. BOX genes were shown to have the highest discrimination compared with ribotyping, PCR fingerprinting, PFGE and restriction fragment end labeling [53]. However BOX-fingerprinting was quite laborious and required internal markers to ease comparison.

Finally, even further specific characterization techniques include penicillin-binding protein typing [79] and Pneumococcal surface protein A (PspA) typing [76,77]. It is important to note that these are not genotyping methods *per se*, but are methods to further understand penicillin resistance and antigenicity of *S. pneumoniae*. Penicillin-binding protein typing investigates the recombination of the penicillin-resistant gene, which became popular after the discovery of penicillin-resistant *S. pneumoniae* [79]. On the other hand, PspA typing has shown that pneumococcal isolates can be grouped into at least six different clades [80]. PspA is an important virulence factor that is a vaccine candidate [81].

Conclusion & future perspective

Genotyping methods used for typing *S. pneumoniae* are important to study the relatedness of isolates, transmission routes and sources of infectious diseases and assess the impact of human interventions or disease outcomes. Ideally, genotyping methods should be adopted in conjunction with the traditional serotyping method so as to gain further genetic information of the invasive strains and pneumococcal isolates causing disease. The pneumococcus continues to travel globally through both carriage and infectious disease such as pneumonia. Techniques with high-throughput, discrimination, typeability and low costs are deemed favorable. With a variety of different techniques and protocols, it has been shown that MLVA is emerging as a highly discriminatory, quick and inexpensive method compared with the

Table 3. The Hunter-Gaston Diversity of multilocus variable number of tandem repeat analysis loci published from different papers.

Locus from *Streptococcus pneumoniae*	Koeck *et al.* [20]	Yaro *et al.* [60]	Pichon *et al.* [21]	Elberse *et al.* [67]	Van Cuyck *et al.* [22]	Rakov *et al.* [23]
Spneu15 BOX-03	0.827	†		0.779	0.607	
Spneu17	0.883	†	†		0.852	0.341
Spneu19	0.749		†		0.674	
Spneu25 BOX-13	0.744	†	†	0.817	0.788	0.514
Spneu26	0.688	†			0.714	
Spneu27	0.575	†			0.561	0.682
Spneu31	0.763	†			0.695	0.139
Spneu32 BOX-02	0.629	†		0.685	0.598	
Spneu33 BOX-04	0.858	†		0.838	0.737	
Spneu34	0.598	†	†		0.682	0.470
Spneu35	0.561	†			0.572	
Spneu36	0.866		†		0.793	
Spneu37 BOX-12	0.876	†	†	0.817	0.855	
Spneu38 BOX-06	0.616	†		0.700	0.557	
Spneu39	0.862	†	†		0.812	0.754
Spneu40 BOX-01	0.832	†		0.829	0.789	0.846
Spneu41	0.484	†			0.567	
Spneu42	0.739					
BOX-11				0.460		
B4						0.704
B10						0.634
B12						0.766

†Loci used in published multilocus variable number of tandem repeat analysis technique, however, no diversity index has been provided.

currently universally used MLST. However further study is required to develop a universal MLVA technique since current protocols have limitations such as unamplified loci. Unless a universally accepted method is developed, comparisons of pneumococcal MLVA types between countries will be difficult. Meanwhile, MLST remains as the 'golden standard' genotyping method for *S. pneumoniae*.

Advances in future genotyping techniques may allow the combination of many genes, including BOX loci and housekeeping genes, with MS or NGS, decreasing cost and time. Both MS and NGS have been used as a platform for MLST. The current method of choice for genotyping *S. pneumoniae* seems to be MLST, and investigations that aim to decrease time and money surely put it in favor for future epidemiology studies. This may imply that MLVA would not become a big prospect in the future, especially since it is not the recent method of choice due to a lack of a universal method. However MLVA has been shown to be cheaper, faster, less demanding and significantly more discriminatory than MLST, which are significant advantages especially when genotyping becomes routine in diagnostic laboratories.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

EXECUTIVE SUMMARY

- There are a variety of characterization techniques for *Streptococcus pneumoniae* including serotyping, multilocus sequence typing (MLST), pulsed field gel electrophoresis, multilocus variable number of tandem repeat analysis (MLVA), next-generation sequencing (NGS) and mass spectrometry (MS).

Traditional characterization methods

- Quellung serotyping, first developed in 1902, is the 'gold standard' characterization method of *S. pneumoniae*.

MLST & PFGE: current 'gold standard' genotyping methods

- Genotyping *S. pneumoniae* with MLST was developed in 1998 and has since become a universally used technique, recognized as a 'gold standard' genotyping method.

MLST & NGS

- With advancing technology, MLST can be combined with NGS to quickly process samples.
- A barcode can be assigned to bacterial species, enabling analysis of multiple bacteria at the same time.

MLST & MS

- MLST has also been combined with MS as an alternative technique to quickly process samples.

MLVA: the emerging genotyping method

- An alternative bacterial fingerprinting method based on variable number of tandem repeats instead of housekeeping genes, MLVA was developed in 2005 for *S. pneumoniae*.
- A number of different MLVA protocols exist, and further investigation and modification continue to devise a faster, cheaper and highly discriminatory method compared with the 'gold standard' MLST.

Feasibility of applying MLVA

- MLVA has demonstrated that it is highly discriminatory, faster and cheaper than MLST, however, modifications still continue as a robust method has not been universally applied that can ensure complete bacterial profiles are achieved.

Other *S. pneumoniae* characterization techniques

- BOX-fingerprinting, penicillin-binding protein typing and PspA typing have been developed as alternative or additional characterization methods of *S. pneumoniae*.

Conclusion & future perspective

- Genotyping methods can be applied in conjunction with traditional serotyping methods.
- MLVA has been shown to be highly discriminatory, cheaper and faster than the 'gold standard' MLST and is therefore ideal for epidemiology studies of *S. pneumoniae*. Unfortunately, further investigation is required to determine a robust and universal MLVA method.
- Genotyping *S. pneumoniae* is predicted to become a common and routine process in diagnostic and clinical laboratories, not only research laboratories as sequencing techniques become faster and cheaper. Currently, MLST is at the forefront of *S. pneumoniae* genotyping, however MLVA is emerging as a possible alternative.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 O'Brien K.L., Wolfson L.J., Watt J.P. *et al.* Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* 374(9693), 893–902 (2009).

2 Weinberger D.M., Malley R., Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. *Lancet* 378(9807), 1962–1973 (2011).

3 Mehr S., Wood N. *Streptococcus pneumonia* – a review of carriage, infection, serotype replacement and vaccination. *Paediat. Resp. R.* 13(4), 258–264 (2012).

4 Brueggemann A.B., Pai R., Crook D.W., Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog.* 3(11), e168 (2007).

5 Golubchik T., Brueggemann A.B., Street T. *et al.* Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat. Genet.* 44(3), 352–356 (2012).

6 Wyres K.L., Klugman K.P., Parkhill J. *et al.* Pneumococcal capsular switching: a historical

perspective. *J. Infect. Dis.* 207(3), 439–449 (2013).

• **Review on pneumococcal capsule switching, emphasizing the impacts since vaccine introduction.**

7 Wang X, Huang B, Blair B, Eglezos S, Bates J. Selection of optimal combinations of loci by the Optimal Combination Finder computer program from a group of variable number tandem repeat loci for use in *Staphylococcus aureus* food poisoning case investigations. *J. Med. Microbiol.* 61(5), 631 (2012).

8 Elberse KEM, Nunes S, Sa-Leao R, Van Der Heide HGJ, Schouls LM. Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. *PLoS ONE* 6(5), e20390 (2011).

•• **Development of an eight-loci multilocus variable number of tandem repeat analysis (MLVA) method for genotyping *Streptococcus pneumoniae*.**

9 Neufeld F. Ueber die Agglutination der Pneumokokken und uber die Theorien der Agglutination. *Zeitschrift fur Hygiene Infektionskrankheiten* 54–72 (1902).

• **Original Quellung serotyping method developed for characterizing the pneumococcal capsules.**

10 Hermans PW, Sluijter M, Hoogenboezem T, Heersma H, Van Belkum A, De Groot R. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. *J. Clin. Microbiol.* 33(6), 1606–1612 (1995).

11 Bentley SD, Aanensen DM, Mavroidi A *et al.* Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* 2(3), e31 (2006).

12 Ko KS, Baek JY, Song JH. Capsular gene sequences and genotypes of "Serotype 6E" *Streptococcus pneumoniae* isolates. *J. Clin. Microbiol.* 51(10), 3395–3399 (2013).

13 Calix JJ, Nahm MH. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *J. Infect. Dis.* 202(1), 29–38 (2010).

14 Calix JJ, Porambo RJ, Brady AM *et al.* Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *J. Biolog. Chem.* 287(33), 27885 (2012).

15 Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MCC, Nahm MH. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* 45(4), 1225–1233 (2007).

16 Bratcher PE, Kim K, Kang JH, Hong JY, Nahm MH. Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. *Microbiology* 156(2), 555–560 (2010).

17 Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 144(11), 3049–3060 (1998).

•• **'Gold standard' multilocus sequence typing genotyping method for *S. pneumoniae* developed.**

18 Lefevre JC, Faucon G, Sicard AM, Gasc AM. DNA fingerprinting of *Streptococcus pneumoniae* strains by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 31(10), 2724–2728 (1993).

19 Donkor ES. Molecular typing of the *Pneumococcus* and its application in epidemiology in sub-Saharan Africa. *Front. Cell. Infect. Microbiol.* 3 (7), 1–5 (2013).

20 Koeck J, Nanjanpop-Lafourcade B, Cade S *et al.* Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. *BMC Microbiol.* 5(1), 66–74 (2005).

•• **First MLVA method developed for genotyping *S. pneumoniae*, and uses 16 loci.**

21 Pichon B, Moyce L, Sheppard C *et al.* Molecular typing of pneumococci for investigation of linked cases of invasive pneumococcal disease. *J. Clin. Microbiol.* 48(5), 1926–1928 (2010).

22 Van Cuyck H, Pichon B, Leroy P *et al.* Multiple-locus variable-number tandem-repeat analysis of *Streptococcus pneumoniae* and comparison with multiple loci sequence typing. *BMC Microbiol.* 12, 241–258 (2012).

•• **Updated method of Koeck *et al.* (2005) MLVA method, reducing the number of loci to seven instead of 16.**

23 Rakov AV, Ubukata K, Ashley Robinson D. Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus *boxB* sequence typing. *J. Mol. Epidemiol. Evolut. Genet. Infect. Dis.* 11(8), 1929–1939 (2011).

24 Chiou CS, Hung CS, Torpdahl M *et al.* Development and evaluation of multilocus variable number tandem repeat analysis for fine typing and phylogenetic analysis of *Salmonella enterica* serovar Typhimurium. *Int. J. Food Microbiol.* 142(1), 67–73 (2010).

25 Larsson JT, Torpdahl M, Nielsen EM, M.W. Grp. proof-of-concept study for successful inter-laboratory comparison of MLVA results. *Euro Surveill.* 18(35), 20–27 (2013).

26 Dunne EM, Ong EK, Moser RJ *et al.* Multilocus sequence typing of *Streptococcus pneumoniae* by use of mass spectrometry. *J. Clin. Microbiol.* 49(11), 3756–3760 (2011).

27 Massire C, Ranken R, Blyn LB *et al.* Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. *J. Clin. Microbiol.* 50(6), 2018 (2012).

28 Boers SA, Van Der Reijden WA, Jansen R. High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS ONE* 7(7), e39630 (2012).

29 Singh P, Foley SL, Nayak R, Kwon YM. Multilocus sequence typing of *Salmonella* strains by high-throughput sequencing of selectively amplified target genes. *J. Microbiol. Methods* 88(1), 127–133 (2012).

30 Hyytiä-Trees EK, Cooper K, Ribot EM, Gerner-Smidt P. Recent developments and future prospects in subtyping of foodborne bacterial pathogens. *Future Microbiol.* 2(2), 175–185 (2007).

31 Coskun-Ari FF, Guldemir D, Durmaz R. One-step multiplex PCR assay for detecting *Streptococcus pneumoniae* serogroups/types covered by 13-valent pneumococcal conjugate vaccine (PCV13). *PLoS ONE* 7(12), e50406 (2012).

32 Carvalho MDG, Hollingshead SK, Pilishvili T *et al.* PCR-based quantitation and clonal diversity of the current prevalent invasive serogroup 6 pneumococcal serotype, 6C, in the United States in 1999 and 2006 to 2007. *J. Clin. Microbiol.* 47(3), 554–559 (2009).

33 Siira L, Kaijalainen T, Lambertsen L, Nahm MH, Toropainen M, Virolainen A. From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. *J. Clin. Microbiol.* 50(8), 2727–2731 (2012).

34 Miernyk K, Debyle C, Harker-Jones M *et al.* Serotyping of *Streptococcus pneumoniae* isolates from nasopharyngeal samples: use of an algorithm combining microbiologic, serologic, and sequential multiplex PCR techniques. *J. Clin. Microbiol.* 49(9), 3209–3214 (2011).

35 Marchese A, Esposito S, Coppo E *et al.* Detection of *Streptococcus pneumoniae* and identification of pneumococcal serotypes by Real-Time polymerase chain reaction using blood samples from Italian children ≤5 years

of age with community-acquired pneumonia. *Microbiol. Drug Res.* 17(3), 419–424 (2011).

36 Jourdain S, Dreze P-A, Vandeven J, Verhaegen J, Van Melderen L, Smeesters PR. Sequential multiplex PCR assay for determining capsular serotypes of colonizing *S. pneumoniae*. *BMC Infect. Dis.* 11(1), 100–100 (2011).

37 Elberse K, Witteveen S, Van Der Heide H *et al.* Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. *PLoS ONE* 6(9), e25018 (2011).

38 Njanpop Lafourcade B-M, Sanou O, Van Der Linden M *et al.* Serotyping pneumococcal meningitis cases in the African meningitis belt by use of multiplex PCR with cerebrospinal fluid. *J. Clin. Microbiol.* 48(2), 612–614 (2010).

39 Azzari C, Moriondo M, Indolfi G *et al.* Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. *PLoS ONE* 5(2), e9282 (2010).

40 Moore CE, Peacock SJ, Sengduangphachanh A *et al.* Enhanced determination of *Streptococcus pneumoniae* serotypes associated with invasive disease in Laos by using a real-time polymerase chain reaction serotyping assay with cerebrospinal fluid. *Am. J. Trop. Med. Hygiene* 83(3), 451–457 (2010).

41 Iraurgui P, Torres MJ, Gandia A *et al.* Modified sequential multiplex PCR for determining capsular serotypes of invasive pneumococci recovered from Seville. *Clin. Microbiol. Infect.* 16(9), 1504–1507 (2010).

42 Saha SK, Luby SP, Santosham M *et al.* Identification of serotype in culture negative pneumococcal meningitis using sequential multiplex PCR: implication for surveillance and vaccine design. *PLoS ONE* 3(10), e3576 (2008).

43 Tarrago D, Fenoll A, Sanchez-Tatay D *et al.* Identification of pneumococcal serotypes from culture-negative clinical specimens by novel real-time PCR. *Clin. Microbiol. Infect.* 14(9), 828–828 (2008).

44 Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *J. Clin. Microbiol.* 44(1), 124–131 (2006).

45 Pneumococcal Molecular Epidemiology Network. Pneumococcal Molecular Epidemiology Network (PMEN) – homepage. http://web1.sph.emory.edu/PMEN/

• **This publication made use of the *Streptococcus pneumoniae* MLST website (http://pubmlst.org/spneumoniae/) sited at the University of Oxford (Jolley & Maiden 2010, BMC Bioinformatics, 11:595). The development of this site has been funded by the Wellcome Trust.**

46 PubMLST - Public databases for molecular typing and microbial genome diversity. http://pubmlst.org

47 Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr. Opin. Microbiol.* 2(3), 312–316 (1999).

48 Schouls L, Van der Heide H. *Streptococcus pneumoniae*. www.mlva.net/spneumoniae/default.asp

49 Enright MC, Knox K, Griffiths D, Crook DWM, Spratt BG. Molecular typing of bacteria directly from cerebrospinal fluid. *Eur. J. Clin. Microbiol. Infect. Dis.* 19(8), 627–630 (2000).

50 Feil EJ, Enright MC, Spratt BG. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Res. Microbiol.* 151(6), 465–469 (2000).

51 Malachowa N, Sabat A, Gniadkowski M *et al.* Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, *spa* typing, and multilocus sequence typing for clonal characterization of Staphylococcus aureus isolates. *J. Clin. Microbiol.* 43(7), 3095–3100 (2005).

52 Jefferies J, Clarke SC, Diggle MA, Smith A, Dowson C, Mitchell T. Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. *Mol. Biotechnol.* 24(3), 303–307 (2003).

53 Hall LMC. Application of molecular typing to the epidemiology of *Streptococcus pneumoniae*. *J. Clin. Pathol.* 51, 270–274 (1998).

54 Tenover FC, Arbeit RD, Goering RV *et al.* Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microbiol.* 33(9), 2233–2239 (1995).

55 Lindstedt B. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 26(13), 2567–2582 (2005).

56 Crisafulli G, Guidotti S, Muzzi A *et al.* An extended multi-locus molecular typing schema for *Streptococcus pneumoniae* demonstrates that a limited number of capsular switch events is responsible for serotype heterogeneity of closely related strains from different countries. *Infect. Genet. Evol.* 13, 151–161 (2013).

57 Moschioni M, Lo Sapio M, Crisafulli G *et al.* Sequence analysis of 96 genomic regions identifies distinct evolutionary lineages within CC156, the largest *Streptococcus pneumoniae* clonal complex in the MLST database. *PLoS ONE* 8(4), e61003 (2013).

58 Van Belkum A, Welker M, Erhard M, Chatellier S. Biomedical mass spectrometry in today's and tomorrow's clinical microbiology laboratories. *J. Clin. Microbiol.* 50(5), 1513–1517 (2012).

59 Ikryannikova LN, Greub G, Filimonova AV *et al.* Discrimination between *Streptococcus pneumoniae* and *Streptococcus mitis* based on sorting of their MALDI mass spectra. *Clin. Microbiol. Infect.* 19(11), 1066–1071 (2013).

60 Williamson YM, Moura H, Woolfitt AR *et al.* Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 74(19), 5891–5897 (2008).

61 Van Ert MN, Schupp JM, Simonson TS *et al.* Mass spectrometry provides accurate characterization of two genetic marker types in *Bacillus anthracis*. *BioTechniques* 37(4), 642–651 (2004).

62 Yaro S, Lourd M, Traore Y *et al.* Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. *Clin. Infect. Dis.* 43(6), 693–700 (2006).

63 Croucher NJ, Vernikos GS, Parkhill J, Bentley SD. Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics* 12(1), 120–120 (2011).

64 Knutsen E, Johnsborg O, Quentin Y, Claverys J-P, Havarstein LS. BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. *J. Bacteriol.* 188(23), 8307–8312 (2006).

65 Multiple-Locus Variable number tandem repeat Analysis. www.mlva.net

66 Martin B, Humbert O, Camara M *et al.* A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res.* 20(13), 3479–3483 (1992).

67 Koeuth T, Versalovic J, Lupski JR. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Res.* 5(4), 408–418 (1995).

68 Van Belkum A, Sluijuter M, De Groot R, Verbrugh H, Hermans PW. Novel BOX

repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. *J. Clin. Microbiol.* 34(5), 1176–1179 (1996).

69 Elberse KEM, Van De Pol I, Witteveen S *et al.* Population structure of invasive *Streptococcus pneumoniae* in The Netherlands in the pre-vaccination era assessed by MLVA and capsular sequence typing. *PLoS ONE* 6(5), e20390 (2011).

70 Chiou CS, Izumiya H, Thong KL *et al.* A simple approach to obtain comparable *Shigella sonnei* MLVA results across laboratories. *Int. J. Med. Microbiol.* 303(8): 678–684 (2013).

71 Multi Locus Sequence Typing. www.mlst.net

72 Database view. http://mlva.u-psud.fr/mlvav4/genotyping

73 Aanensen DM, Spratt BG. The multilocus sequence typing network: mlst.net. *Nucleic Acids Res.* 33, 728–733 (2005).

74 Elberse K. *Streptococcus pneumoniae*. www.mlva.net/spneumoniae/default.asp

75 Koeck JL, Brunetaud J, Koeck JD. MLVA Bacterial Genotyping – home. www.mlva.eu/index.php?lg=2.

76 Elberse KEM, Van Der Heide HGJ, Witteveen S *et al.* Changes in the composition of the pneumococcal population and in IPD incidence in The Netherlands after the implementation of the 7-valent pneumococcal conjugate vaccine. *Vaccine* 30(52), 7644–7651 (2012).

77 O'Sullivan MV, Sintchenko V, Gilbert GL.. Software for selecting the most informative sets of genomic loci for multi-target microbial typing. *BMC Bioinformatics* 14,148 (2013).

78 Robertson GA, Thiruvenkataswamy V, Shilling H *et al.* Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. *J. Med. Microbiol.* 53(Pt 1), 35–45 (2004).

79 Munoz R, Musser JM, Crain M *et al.* Geographic distribution of penicillin-resistant clones of *Streptococcus pneumoniae*: characterization by penicillin-binding protein profile, surface protein A typing, and multilocus enzyme analysis. *Clin. Infect. Dis.* 15(1), 112–118 (1992).

80 Qian J, Liu L, Li C *et al.* Diversity of pneumococcal surface protein A (PspA) and relation to sequence typing in *Streptococcus pneumoniae* causing invasive disease in Chinese children. *Eur. J. Clin. Microbiol. Infect. Dis.* 31(3), 217–223 (2012).

81 Vadesilho CFM, Ferreira DM, Moreno AT *et al.* Characterization of the antibody response elicited by immunization with pneumococcal surface protein A (PspA) as recombinant protein or DNA vaccine and analysis of protection against an intranasal lethal challenge with *Streptococcus pneumoniae*. *Microb. Pathog.* 53(5–6), 243–249 (2012).

Appendix B9: Statement of Contribution for manuscript "Modified MLVA for genotyping Queensland invasive *Streptococcus pneumoniae*" (April 2015)

QUT

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Best viewed in Adobe Reader

Statement of Contribution of Co-Authors for Thesis by Published Paper

The following is the format for the required declaration provided at the start of any thesis chapter which includes a co-authored publication.

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter: Chapter 4

Publication title and date of publication or status: Modified MLVA for genotyping Queensland Invasive *Streptococcus pneumoniae*

Published in: PLoS One, 2015. Volume 10, issue 4, e0121870. doi:10.1371/journal.pone.0121870.

Contributor	Statement of contribution*
Rachael E. Rayner Signature Date 24 July 2015	Wrote the manuscript, contributed to experimental design and research plan, conducted experiments and analysed the data.
John Saul 28/7/015	Contributed to the formulation of the research plan, assisted with writing and reviewed the manuscript, contributed to experimental design, execution and data analysis.
Louise M. Hafner 29/7/15	Contributed to research plan, assisted with writing and reviewed the manuscript.
Flavia Huygens	Formulated the research plan, assisted with writing and reviewed the manuscript, contributed to experimental design and data analysis.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Flavia Huygens
Name

Signature

29/7/15
Date

RSC, Level 4, 88 Musk Ave, Kelvin Grove Qld 4059
Current @ 24/07/2015

Page 1 of 1
CRICOS No. 00213J

Appendix B10: Published manuscript "Modified MLVA for genotyping Queensland invasive *Streptococcus pneumoniae*" to PLoS One (April 2015)

PLOS | ONE

RESEARCH ARTICLE

Modified MLVA for Genotyping Queensland Invasive *Streptococcus pneumoniae*

Rachael E. Rayner[1,2], John Savill[3], Louise M. Hafner[1,2], Flavia Huygens[1,2]*

1 School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia, 2 Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia, 3 Public Health Microbiology Laboratory, Queensland Health Forensic and Scientific Services, Brisbane, Australia

* f.huygens@qut.edu.au

Abstract

Background

Globally, over 800 000 children under five die each year from infectious diseases caused by *Streptococcus pneumoniae*. To understand genetic relatedness between isolates, study transmission routes, assess the impact of human interventions e.g. vaccines, and determine infection sources, genotyping methods are required. The 'gold standard' genotyping method, Multi-Locus Sequence Typing (MLST), is useful for long-term and global studies. Another genotyping method, Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA), has emerged as a more discriminatory, inexpensive and faster technique; however there is no universally accepted method and it is currently suitable for short-term and localised epidemiology studies. Currently Australia has no national MLST database, nor has it adopted any MLVA method for short-term or localised studies. This study aims to improve *S. pneumoniae* genotyping methods by modifying the existing MLVA techniques to be more discriminatory, faster, cheaper and technically less demanding than previously published MLVA methods and MLST.

Methods

Four different MLVA protocols, including a modified method, were applied to 317 isolates of serotyped invasive *S. pneumoniae* isolated from sterile body sites of Queensland children under 15 years from 2007–2012. MLST was applied to 202 isolates for comparison.

Results

The modified MLVA4 is significantly more discriminatory than the 'gold standard' MLST method. MLVA4 has similar discrimination compared to other MLVA techniques in this study). The failure to amplify particular loci in previous MLVA methods were minimised in MLVA4. Failure to amplify BOX-13 and Spneu19 were found to be serotype specific.

Conclusion

We have modified a highly discriminatory MLVA technique for genotyping Queensland invasive *S. pneumoniae*. MLVA4 has the ability to enhance our understanding of the pneumococcal

Citation: Rayner RE, Savill J, Hafner LM, Huygens F (2015) Modified MLVA for Genotyping Queensland Invasive *Streptococcus pneumoniae*. PLoS ONE 10 (4): e0121870. doi:10.1371/journal.pone.0121870

Academic Editor: Baochuan Lin, Naval Research Laboratory, UNITED STATES

Received: July 20, 2014

Accepted: February 20, 2015

Published: April 29, 2015

Data Availability Statement: All relevant data are within the paper and its supporting information files.

Funding: This work was supported by the Queensland Health Forensic and Scientific Services, Coopers Plains, Australia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

epidemiology and the changing genetics of the pneumococcus in localised and short-term studies.

Introduction

More than 800 000 children under five succumb to invasive pneumococcal diseases (IPD) each year globally [1]. Australia is not exempt from invasion with 6.7 per 100 000 notifications in 2013 [2–3]. IPD are defined as the isolation of *S. pneumoniae* from normally sterile body sites including blood, tissues, and cerebrospinal, joint, pericardial or pleural fluids [4]. The changing pneumococcal population structure worldwide is largely the result of serotype replacement and capsule switching, especially in first world countries such as Australia where vaccines are widely implemented [5–7]. Serotype replacement has been problematic as current pneumococcal vaccines only target 13 out of 95 pneumococcal serotypes [8–13]. Capsule switching, the transfer of a capsule genes from one pneumococcus to another, is a regular occurrence in pneumococcal populations however it has the potential to reduce vaccine efficacy because vaccine escape isolates can emerge [6–7]. Vaccine escape isolates can develop within 2–3 years of a vaccine introduction, as already detected in the USA and Italy [6, 14]. *In vitro* studies have demonstrated that capsule switching can also impact pneumococcal virulence properties, particularly since the polysaccharide capsule is a virulence factor of pneumococci [8]. A highly virulent capsule type 5 strain was rendered avirulent when expressing a capsule type 3 and a type 6A strain expressing a capsule type 6C was more virulent than wild types [15–16].

So far, there have been little published studies examining the pneumococcal population structure in Australia since the introduction of the new vaccine 13-valent pneumococcal conjugate vaccine in July 2011 [17].

For decades, the main technique for surveying serotype distribution and replacement has involved serotyping pneumococci with antisera (such as the Quellung reaction) [18]. However serotyping is expensive, laborious, and ambiguous, revealing no information about genetic recombination and capsule switching. Therefore several genotyping methods have been developed, including MLST [19] and several MLVA techniques [20–23]. It is important to bear in mind that genotyping and serotyping are both required in combination to detect capsule switching events.

Commonly, MLST is used to genotype *S. pneumoniae* based on the original technique involving housekeeping genes developed by Enright and Spratt [19]. MLST housekeeping genes are considered to be stable and less prone to recombination than the rest of the genome, enabling examination of long-term population changes and studies across wider areas [19]. However, MLST is expensive and laborious, not suitable for large-scale genotyping or routine use [20–21, 24–27]. There is no national Australian MLST database, and the international MLST database only contains 138 Australian isolates from the year 1967 to 2013, providing very little information regarding the Australian pneumococcal population. The international MLST database could be used as a comparison against the pneumococcal isolates found in the rest of the world.

To reduce cost and labour intensity, MLVA was developed for genotyping *S. pneumoniae*. It is reported to be more discriminatory, inexpensive and faster than MLST, however more suitable for short-term epidemiology changes and localised outbreaks [20–21, 27]. MLVA involves amplifying Variable Number of Tandem Repeats (VNTR) loci and sizing the PCR products according to fragment lengths. VNTRs are suitable genotyping targets since there are multiple

loci throughout the genome and they are polymorphic [28]. MLVA can be multiplexed, and uses a DNA sequencer for high throughput analysis of fragment sizes, not sequencing [25].

Different MLVA protocols exist including Koeck's *et al.* [21] MLVA which amplifies seventeen VNTRs, each in a singleplex PCR. This method was used to genotype pneumococcal isolates in Burkina Faso [29]. Unfortunately the practicality of typing seventeen targets and the difficulty comparing analysis of gels are some of the limitations. There is a freely accessible database (http://mlva.u-psud.fr/mlvav4/genotyping/index.php - Simple Databases labelled 'Streptococcus pneumoniae2005) which allows comparison of profiles, however there are only 59 isolates available in this database. Recently this protocol has been modified by Van Cuyck *et al.* [22] by reducing the number of VNTRs to seven, which were experimentally determined to have the highest discrimination (Hunter-Gaston Diversity >0.8) within a population of 331 UK isolates. This seven marker MLVA is claimed to be a minimum universal set, ideal for genotyping pneumococcal isolates. However it is known that pneumococcal populations differ between countries, therefore the selection of these seven MLVA markers may not be suitable for Australian isolates.

Another MLVA protocol was developed by Elberse *et al.* [20], which utilises eight BOX VNTR loci amplified in two multiplex PCRs with fluorescently labelled probes. BOX loci are a type of VNTR loci, distinguishable by varying numbers of *boxB* repeat regions (45 base-pair) found between b*oxA* and *boxC* which remain stable under laboratory conditions [20, 30]. BOX elements can form secondary structures and can affect the expression of downstream genes [31]. Elberse's method solely uses BOX loci (no other VNTR), while Koeck's method uses a combination of BOX loci and other VNTR loci, which do not contain stable *boxA* and *boxC* units, and vary in repeat lengths e.g. 60 base-pair. Elberse's MLVA protocol has been applied to genotype pneumococci in the Netherlands and carriage isolates from Portugal, as well as tracking a localised outbreak in England [32]. However, a limitation is that some BOX loci fail to amplify (assigned '99') therefore leaving profiles incomplete, an issue that remains unresolved. An MLVA type can still be assigned even if the profile contains a non-amplified locus (studies from Elberse identify that 89% of serotype 7F isolates will commonly have a BOX-06 that won't amplify). The limitation of having an incomplete genetic profile is that we can't see the 'true' bacterial fingerprint of the isolate. It is unknown whether this could have implications for population studies.

Finally, Multi-Locus *boxB* Typing (MLBT) was developed, a variation of MLVA by sequencing VNTR loci to detect Single Nucleotide Polymorphisms (SNPs) as well as fragment length variations [23]. MLBT contains VNTRs that have been used in the other MLVA protocols, however the complexities of MLBT does not enable ease of comparison with other MLVA methods [20–21, 23].

The aim of this study was to improve pneumococcal genotyping methods by modifying existing MLVA. It is important to modify the existing MLVA methods to be more discriminatory, faster, cheaper and technically less demanding than previously published MLVA methods and MLST.

Methods

Setting

All pneumococcal isolates from Queensland patients with IPD are required to be submitted to the Public Health Microbiology Laboratory at Queensland Health Forensic and Scientific Services (QHFSS), Brisbane. Serotyping (Quellung) is mandatory, however further genotyping has only been performed for research purposes. Invasive *S. pneumoniae* isolates taken from

normally sterile body sites were serotyped by the Pneumococcal Reference Laboratory, QHFSS, using Quellung reaction [4, 18].

Laboratory methods

S. pneumoniae isolates were cultured 16-streak on Horse Blood Agar (HBA, Oxoid, Australia—a commercially available product routinely used for the culture of bacteria). A single colony was re-cultured on fresh HBA to ensure isolates were genetically identical. Isolates were boiled in 400μL TE buffer (~pH8.0) for eight minutes to prepare a thermolysate with template DNA and were stored at -20°C until further use.

MLVA markers were selected from previously published papers [20–23]. All MLVA markers were analysed *in silico* to determine the expected fragment sizes and copy numbers against known *S. pneumoniae* genomes from the NCBI database (R6, G54, CGSP14, TIGR4 and Hungary19A-6).

S. pneumoniae isolates (n = 317) detected in Queensland were genotyped using MLVA1, MLVA2 and MLVA4. Elberse's MLVA1 contained two multiplexes with eight BOX genes (Table 1) [20]. The MLVA2 procedure was based on Van Cuyck's MLVA method and contained seven VNTRs (Table 1) [22]. A single multiplex reaction was performed with Spneu17, Spneu19, Spneu27 and Spneu39 as the other three had already been typed in MLVA1. Spneu31 [21] and B10 [23] were separately amplified to determine whether they provided high discrimination within the pneumococcal population, and therefore suitable for the modified MLVA method.

MLVA4 was the modified MLVA method based on the high discrimination of the fourteen loci previously used, and the ability to amplify all loci across the different serotypes detected in Queensland [20–23]. Three multiplexes were developed (Table 1). Four primer sets were redesigned due to difficulties in amplification, including BOX-12 (F: GAGATTGCCCTTTTCATCTTCG; R: AGCAACCATTGAAACGCCTG), BOX-13 (F: TCAAAAGATTGGAGAGTTCCGC; R: GGATTTGGAGAGCAAGCAGATC), Spneu19 (F: CACTCACCGTTAGCATTGACTCG; R: TAATCAGGGAGTAGTTGGTTGGG) and B10 (F: GGAGCCGAGTAGGAGATTCTCAC; R: TCGTAGGCTGCTACATTGACCAG) (Geneworks, Australia).

The Corbett Cas1200 Robotic system was used to prepare mastermix with DNA thermolysate. The PCR protocol was optimised and consisted of 15min at 95°C, 30 cycles of 95°C for 30sec, 58°C for 60sec and 72°C for 60sec, followed by extension of 72°C for 10min and held at 4°C. Diluted PCR products (1:150 reverse-osmosis water (PALL, Australia)) were combined with 1200LIZ internal ladder (Applied Biosystems, Australia). A heat denaturation step (95°C for 5min, followed by a hold step at 4°C) was performed on a thermocycler to separate the dsDNA. Fragment sizing was performed on AB3130 (Applied Biosystems, Australia).

Table 1. Multiplex arrangement for MLVA methods.

Genotyping method	Multiplex number	Loci	Reference
MLVA1	1	BOX-01, BOX-02, BOX-03, BOX-04	[20]
	2	BOX-06, BOX-11, BOX-12, BOX-13	[20]
MLVA2	1	Spneu17, Spneu19, Spneu27, Spneu39	[22]
	2	Spneu25, Spneu33, Spneu37	[22]
MLVA4	1	BOX-01, B10, Spneu19, Spneu39	This study
	2	BOX-12, BOX-13	This study
	3	BOX-02, BOX-03, BOX-04, Spneu17	This study

doi:10.1371/journal.pone.0121870.t001

Finally, MLST was also applied to selected isolates (n = 202) for comparison purposes as previously described [19, 33].

Analysis

PeakScanner V1.0 software was used to analyse MLVA results (Applied Biosystems, Australia). A MLVA type (MT) using MLVA1 was assigned to each isolate using the MLVA database (http://www.mlva.net). MT types for MLVA2, and MLVA4 was manually assigned from our own database. Van Cuyck *et al.* [22] did not provide a database or published MLVA types that could be used in the Queensland pneumococcal population. The pneumococcal population structure using all MLVA methods were displayed as eBurst diagrams produced from PHYLOViZ software [34]. Clonal clusters (CC) are defined as two or more isolates that are genetically related and linked by single locus variants (SLV) or double locus variants (DLV). Where an international or larger database is available, Queensland isolates were compared.

MLST results were analysed using ChromasPro software (Technelysium Pty Ltd.) using batch alignment analysis. Allele numbers and sequence types (ST) were assigned to each isolate from the MLST database (www.mlst.net) and displayed as an eBurst diagram using PHYLOViZ software.

The Simpson's Index of Diversity (S) was calculated to compare the discrimination of all genotyping methods (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool). If the 95% confidence intervals (CI) overlap between methods, the hypothesis that the methods have similar discriminatory power cannot be excluded. The Adjusted Wallace coefficient (AW) was used to measure the congruence between typing methods [35–36].

The frequency of non-amplified loci ('99') was compared between each MLVA method to determine whether this was a limitation of a particular method or associated with specific serotypes. Hunter-Gaston Diversity Index (DI) (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl) was used to calculate the genetic diversity of VNTR genes within the Queensland population, as used in previous studies [21–22, 37]. DI is a measure of the variation of the number of repeats at each locus, ranging from 0.0 (no diversity) to 1.0 (complete diversity).

Ethics statement

No human participants were involved directly in this study and hence, human ethics clearance was not required. *S. pneumoniae* isolates routinely cultured from clinical specimens were used and we investigated the epidemiology of the *S. pneumoniae* isolates, changing genotypes and population structure in Queensland.

Results

Comparison of genotyping methods

All 317 *S. pneumoniae* Queensland isolates in this study were assigned an MLVA genotype, covering 35 serotypes detected in Queensland from 2007–2012 (serotypes include 1, 3, 4, 6A, 6B, 6C, 7F, 8, 9N, 9V, 10A, 10F, 11A, 12F, 14, 15B, 15C, 16F, 18A, 18B, 18C, 19A, 19F, 22A, 22F, 23A, 23B, 23F, 24F, 33B, 33F, 34, 35B, 35F, and 38). MLVA4 method has novel aspects as four sets of primers and the three multiplex PCR has been newly designed in this study.

Isolates selected for MLST included 13vPCV serotypes and non-13vPCV serotypes to minimise labour work and costs (excluding those originally in the 7vPCV i.e. serotype 4, 6B, 9V, 14, 18C, 19F and 23F, and serotype 19A). Studies have already shown that 7vPCV serotypes have significantly declined [38–39], so we focused on the examination of these recently targeted or non-targeted serotypes.

Table 2. Calculation of Simpson's Index of Diversity for serotyping (Quellung), MLST, MLVA1, MLVA2 and MLVA4 methods.

Molecular method	Total number of genotypes	Simpson's Index of Diversity (S)	CI (95%)	Average time to genotype 48 isolates (days)	Cost per isolate (PCR to genotype) (AUS$)	Non-amplified loci (%)
Total Number of S. pneumoniae isolates = 202						
Serotyping	29	0.912	0.893–0.930	-	-	N/A
MLST	49	0.936	0.920–0.952	16–20	346.65	N/A
MLVA1	71	0.963	0.953–0.974	3–4	17.10	20.8
MLVA2	97	0.977	0.970–0.985	3–4	16.37	24.3
MLVA4	106	0.978	0.971–0.986	3–4	23.71	12.4
Total Number of S. pneumoniae isolates = 317						
Serotyping	35	0.914	0.897–0.932	-	-	N/A
MLVA1	163	0.984	0.980–0.989	3–4	17.10	20.2
MLVA2	175	0.987	0.983–0.991	3–4	16.37	21.1
MLVA4	203	0.990	0.987–0.994	3–4	23.71	12.6

doi:10.1371/journal.pone.0121870.t002

All three MLVA genotyping methods had a higher discriminatory power compared to MLST (Table 2). MLVA4 had a significantly higher discrimination compared to MLST (S = 0.978 with 106 MLVA types; S = 0.936; 49 ST's) (n = 202). In comparison, MLVA1 had a discrimination of S = 0.963 (n = 202; 71 MLVA types) and MLVA2 had S = 0.977 (n = 202; 97 MLVA types).

However, when comparing the Adjusted Wallace Coefficient of MLVA4 with the other two MLVA methods, all MLVA methods had similar discriminatory power (Table 2). MLVA4 has high congruence with MLVA1 (AW = 0.883), MLVA2 (AW = 0.766) and MLST (AW = 0.966). This indicates that the MLVA4 genotypes will have a 96.6% probability that it will have the same MLST type. Conversely, the congruency of MLST with MLVA4 (AW = 0.314) indicates that the MLST types will have a 31.4% probability of having the same MLVA4 types, indicating that MLVA4 is more discriminatory.

New MLVA1 BOX alleles were also found including BOX-04 fragment sizes 10 repeats (n = 1), 11 repeats (n = 1) and 12 repeats (n = 3), BOX-03 fragment size 8 repeats (n = 2), BOX-12 fragment size 0 repeats (n = 2) and BOX-13 fragment size >2000bp (unsure how many repeats until sequenced; n = 8). There are 107 new MLVA1 MT types not recorded by the MLVA database [25] In addition, 20 new STs have been submitted to the MLST international database (ID#23211–23429).

VNTR loci and '99'

BOX-01, BOX-04, BOX-12, BOX-13 and Spneu17 had the highest diversity (DI ≥ 0.80) while BOX-02, BOX-11, Spneu19 and Spneu27 had the lowest diversity (DI ≤ 0.66) (Table 3). When examining MLVA1, 20.2% of isolates still contain at least one non-amplified locus. Even MLVA2 contained 21.1% of isolates with non-amplified loci. MLVA4 method reduces the '99' to 12.6% (Table 4).

Table 3. Adjusted Wallace Coefficient and 95% confidence intervals (CI) for four genotyping and one serotyping method.

Total number of S. pneumoniae isolates = 202					
	MLVA1	**MLVA2**	**MLVA4**	**Serotyping**	**MLST**
MLVA1		0.524 (0.461–0.5287)	0.512 (0.442–0.582)	0.940 (0.891–0.989)	0.984 (0.974–0.995)
MLVA2	0.871 (0.800–0.942)		0.739 (0.667–0.810)	0.880 (0.804–0.956)	0.916 (0.887–0.945)
MLVA4	0.883 (0.823–0.942)	0.766 (0.690–0.843)		0.918 (0.842–0.994)	0.966 (0.927–1.000)
Serotyping	0.372 (0.301–0.442)	0.209 (0.157–0.262)	0.211 (0.161–0.260)		0.648 (0.587–0.709)
MLST	0.552 (0.456–0.648)	0.309 (0.244–0.374)	0.314 (0.254–0.374)	0.919 (0.869–0.969)	

doi:10.1371/journal.pone.0121870.t003

These VNTR loci failed to amplify in specific serotypes. BOX-06 failed to amplify 75% of serotype 7F (n = 27), BOX-13 failed to amplify 42% of serotype 33F (n = 5) and 54% of serotype 6C (n = 7), Spneu19 failed to amplify 63% of serotype 3 (n = 12) and 50% of serotype 38 (n = 4), and BOX-01 failed to amplify 20% of serotype 19F isolates (n = 3) (Table 3). Spneu19 allele sequences were not found in serotype 3 genomes using BlastN (strains OXC141, SP3, SPN021198, SPN034156, SPN034183, SPN072838, SPN994038 and SPN994039). Similarly Spneu19 failed to amplify in serotype 38; however no serotype 38 genome has been sequenced to date to allow us to determine whether the Spneu19 locus exists. No match was found for BOX-13 in serotype 6C NCBI genomes, and failed to amplify in serotype 33F, however gel electrophoresis showed a large >2000bp fragment (S1 Fig.). BOX-06 primers appear to anneal to a serotype 7F contig (CDC 1087–00 contig ABFT01000005.1) however the fragment size exceeds 1200bp which would not be detected with the AB3130 internal size ladder LIZ1200. Gel electrophoresis did not reveal any bands, corresponding to BOX-06 in serotype 7F.

Comparison of MLVA and MLST eBurst

MLVA4 has been shown to be vastly cheaper and faster than MLST, costing $23.71 compared to $346.65 per isolate, respectively, and taking only 3–4 days on average to genotype 48 isolates, compared to 16–20 days, respectively.

Comparison of MLST and our MLVA4 eBurst analysis clearly shows two different population structures (Fig. 1). The pneumococcal population structure is displayed according to the MLVA4 genotypes, and it is observed that there are 32 clonal clusters (CC), with the larger eleven clusters labelled in the figure. The MLST results are overlayed on the MLVA4 results; therefore each colour represents a different MLST type. It can be seen that several MLVA4 clusters would appear as a single MLST type, for example CC1 which predominantly contains serotype 7F isolates has as many as ten different MLVA4 types however only one MLST type as depicted by the single green colour of the circles.

Using MLVA4 for localised Queensland studies show that most CCs predominantly contain a single serotype (e.g. CC1 contains serotype 7F, CC4 contains serotype 3) (Fig. 1). However some clusters contain mixed serotypes including CC11 (serotype 19A and 19F), CC8 (serotype 15B, 15C and 19A), CC9 (serotype 18B and 18C), CC10 (serotype 1 and 4), CC3 (serotype 6A and 6C), CC12 (serotype 8 and 11A), CC13 (serotype 33F and 22F). MLST identifies most of these CC but many appear as a single isolates.

MLST results were also overlayed with the MLVA1 and MLVA2 results. Both MLVA methods are more discriminatory than MLST, as evident by the diversification of MLST types into many MLVA types (e.g. CC10 contains at least five different MLVA1 types with the same MLST type) (Figs. 2 and 3). Interestingly, MLVA1 groups serotypes 1, 3 and 7F into the same cluster (CC10) according to the genotype profile (Fig. 2). Similarly, MLVA2 groups most of the clusters together as isolates appear to have SLV or DLV of another isolates (Fig. 3). However

Table 4. Hunter-Gaston Diversity (DI) for all selected MLVA loci (n = 317 isolates) and frequency of non-amplified loci '99'.

MLVA method	Locus	Diversity Index (DI)	CI	Size of tandem repeat (bp)	K	Number of non-amplified loci before singleplex	Number of non-amplified loci after singleplex	Non-amplified serotypes (% of isolates)
MLVA2, MLVA4	Spneu17	0.853	0.840–0.866	45	12	6	0	
MLVA1, MLVA2, MLVA4	BOX-13	0.821	0.806–0.836	45	10	23	13	33F (5); 6C (7); 19F (1)
MLVA1, MLVA2, MLVA4	BOX-12	0.805	0.778–0.833	45	13	89	4	3 (3); 23B (1)
MLVA1, MLVA4	BOX-01	0.801	0.785–0.817	45	10	11	4	19F (3); 33B (1)
MLVA1, MLVA4	BOX-04	0.797	0.771–0.823	45	12	7	0	
MLVA2, MLVA4	Spneu39	0.794	0.775–0.812	45	10	0	0	
MLVA1, MLVA4	BOX-03	0.789	0.754–0.824	45	14	28	0	
MLVA4	BOX-10	0.784	0.748–0.821	45	11	54	4	15C (2); 19A (2)
	Spneu31	0.729	0.696–0.762	45	11	7	2	
MLVA1, MLVA2	BOX-06	0.696	0.653–0.739	45	8	53	32	
MLVA2, MLVA4	Spneu19	0.663	0.632–0.694	60	9	24	16	3 (12); 38 (4)
MLVA1, MLVA4	BOX-02	0.651	0.628–0.674	45	5	1	0	
MLVA2	Spneu27	0.637	0.601–0.674	45	8	13	3	
MLVA1	BOX-11	0.392	0.341–0.443	45	3	14	7	
	TOTAL	-	-	-	-	330	121	

'K' represents the number of different repeats present for each locus in the Queensland pneumococcal sample set.

doi:10.1371/journal.pone.0121870.t004

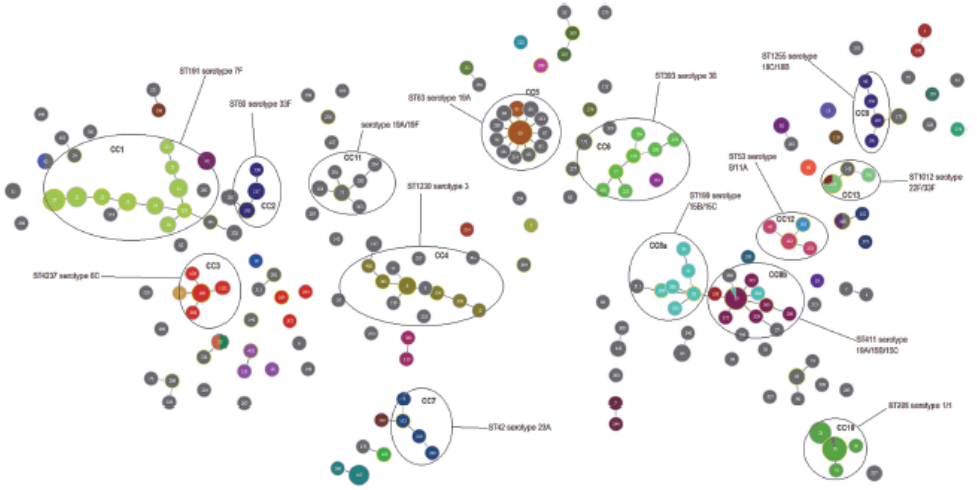


Fig 1. MLVA4 eBurst of Queensland invasive *S. pneumoniae* isolates from 2007 to 2012 (n = 317). MLST eBurst of invasive *S. pneumoniae* isolates from 2007–2012 (n = 202). MLST results are overlayed (coloured circles) to allow ease of comparison against MLVA4 results. Grey circles represent isolates that have not been assigned a MLST type. The size of the isolate circles corresponds to the number of isolates. Clonal clusters are identified in black ovals, and those that contain predominantly one colour indicate that the isolates have a single MLST type. Isolates are labelled with MLVA4 type (MT).

doi:10.1371/journal.pone.0121870.g001

for some serotypes, MLVA4 provides higher discrimination than MLST and the other MLVA methods, evident by the high diversification of serotype 7F and serotype 3 (Fig. 4).

Finally, MLST and MLVA4 used in combination with serotyping can identify potential capsule switching. MLST identifies seven potential capsule switches including serotypes 22F to 33F (ST1012), serotypes 18B to 18C (ST1255), serotypes 15B to 15C (ST199), serotypes 1 to 4 (ST306), serotypes 19A, 15B and 15C (ST411), serotypes 6A to 6C (ST4237) and serotypes 8 to 11A (ST53) (S1 Fig.). On the other hand, MLVA4 only detects two capsule switches including serotypes 15C to 19A (MT59; ST411), and serotypes 1 to 4 (MT36; ST306). The Queensland *S. pneumoniae* isolates are listed in the supplementary data (S1 Table) which contains MLVA1, MLVA4, MLST and serotype for comparison.

Discussion

MLVA has emerged as an alternative genotyping technique to the 'gold standard' MLST as it has higher discriminatory power, is fast and inexpensive [20–21, 24]. Results in this study support this finding as MLVA4 is vastly cheaper and faster than MLST, and is still comparable to the costs and laboratory processing time of the other published MLVA methods (Table 2). Reducing time and costs for genotyping will have an impact on the public health field by having the ability to resolve "large and complex outbreak situations" [24].

The choice of the ten VNTRs for our MLVA4 method was based on a Hunter-Gaston Diversity of $\geq$0.8, as well as an anchor locus with low discrimination to determine long-term changes (BOX-02), and an extra locus with high discriminatory power for specific serotypes (Spneu19). Of the eight highly discriminatory VNTR loci used in previously published MLVA

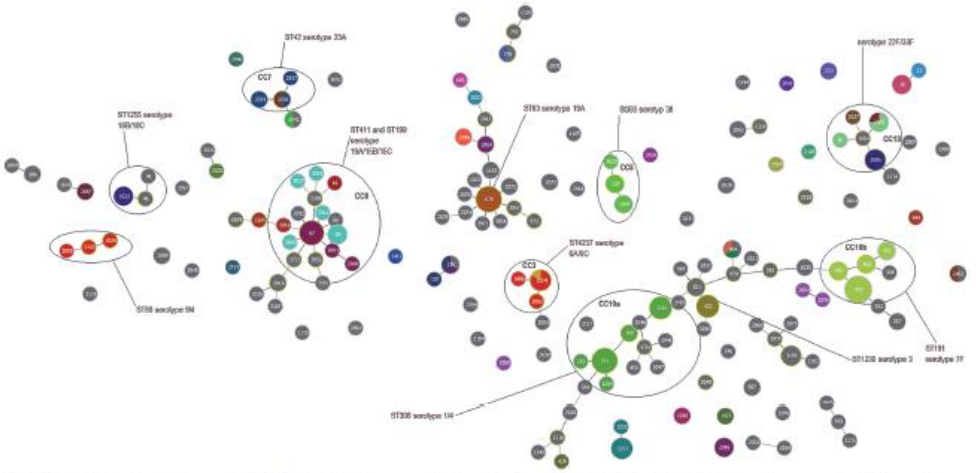


Fig 2. The MLVA1 eBurst diagram depicting invasive *S.pneumoniae* isolates from 2007 to 2012 (n = 317). MLST results are overlayed (coloured circles) to allow ease of comparison against MLVA1 results. Grey circles represent isolates that have not been assigned a MLST type. CC are surrounded by a black oval and contain SLV and DLV. The size of the dots corresponds to the number of isolates.

doi:10.1371/journal.pone.0121870.g002

papers, seven VNTRs are included in the modified MLVA4 method [20–23]. MLVA4 maintains high congruence with the other MLVA methods and MLST (Table 2).

When comparing the different genotyping methods, MLVA4 method maintains a high discriminatory power whilst minimising the number of '99', and is significantly more accurate in representing the Queensland pneumococcal population structure than MLST. Clusters of isolates can be observed in more detail and can glean more information when combined with isolate information such as antibiotic resistance or location of disease (not investigated in this study). Admittedly, additional markers have increased the discriminatory power when applied to Queensland, Australian isolates however we have maintained minimal laboratory work to three multiplexes.

MLST was proven to be less discriminatory although it could still maintain as an ideal method for long-term and international epidemiology studies. A less discriminatory protocol is problematic if it does not detect emerging genotypes or outbreaks in a localised area or state. It is evident that MLVA4 is more efficient at discriminating pneumococcal isolates because, for example, MLST type 191 (serotype 7F) actually can be separated into nine different MLVA4 types (Fig. 4).

However, this study confirms the problem of failing to amplify particular loci, resulting in incomplete genotypes. As a result, a number of primers were re-designed in this study in an attempt to successfully amplify the failing loci (including BOX-10, BOX-12, BOX-13 and Spneu19). Elberse *et al.* [40] reported that 24% of their isolates still contained one or more non-amplified BOX loci even after repeated PCR. This study observed that some loci failed to amplify in specific serotypes, for example BOX-06 failed to amplify 75% of Queensland serotype 7F, indicating that primers needed to be redesigned, or VNTR fragments were absent. Serotype 7F (89% isolates) has been associated with a large number of non-amplified BOX-06

Fig 3. The MLVA2 eBurst diagram depicting invasive S. pneumoniae isolates from 2007 to2012 (n = 317). MLST results are overlayed (coloured circles) to allow ease of comparison against MLVA1 results. Grey circles represent isolates that have not been assigned a MLST type. CC are surrounded by a black oval and contain SLV and DLV. The size of the dots correspond to the number of isolates.

doi:10.1371/journal.pone.0121870.g003

genes [40]. Since the primers for MLVA are located in stable *boxA* and *boxC* regions flanking the *boxB* repeats, the difficulty in amplifying the loci is unlikely. Therefore, it is theorised that gene elements are more likely to be lost (or acquired) if there is a higher average number of *boxB* repeats [31]. It is possible that the BOX-06 region is missing from serotype 7F as there are up to eight different BOX-06 fragment lengths. Therefore BOX-06 was not used in MLVA4 due to a high percentage of '99' and a low Hunter-Gaston diversity. Serotype 7F was the second most common serotype (9%) found in Queensland, therefore using higher discriminatory loci was favoured.

Similarly, Spneu19 loci and BOX-13 could not be detected in serotype 3 and 33F isolates, respectively. The failure to amplify Spneu19 in serotype 3 has also been observed, suggesting that serotype 3 lack *pcpA* which codes for a non-essential surface protein involved in cell adhesion [21, 41]. On the other hand, long BOX-13 fragments (>2000bp) have been identified in serotype 33F isolates, possibly accounting for the '99' results since the AB3130 internal size ladder only reaches 1200bp. Large fragments could be explained by the placement of an insertion sequence (IS) element, making the BOX element appear to be larger than 2kb. The presence of IS elements has been described in another MLVA study [42].

Variations in interpreting the population structure of *S. pneumoniae* have been observed when using different genotyping protocols. Already potential capsule switches have been observed between a serotype 19A and 15C in CC8 (MT59; ST411), and a serotype 1 and 4 (MT36; ST306) in CC10 in our Queensland population using MLVA4 (S1 Fig.). MLST also identifies these capsule switches, as well as many others which may indicate false capsule switching since MLST is less discriminatory to discern true genetic background of *S. pneumoniae* isolates. When examining the international database, some of these capsule switches can be verified, for

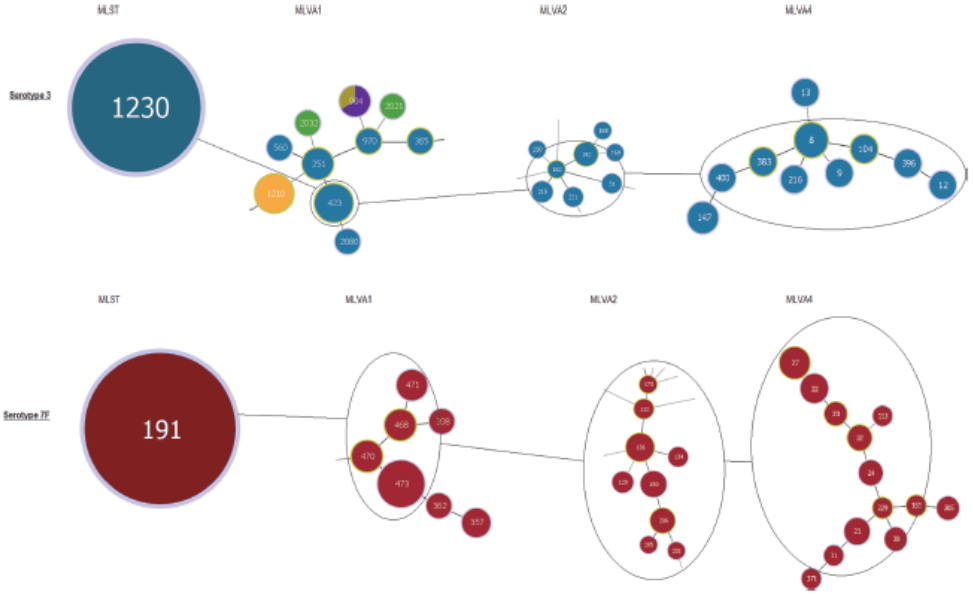


Fig 4. Comparison of discriminatory power for MLST, MLVA1, MLVA2 and MLVA4 when genotyping serotype 3 and serotype 7F in Queensland, Australia. The population structure based on the respective genotypes determined by each genotyping method varies. MLST is considered the least discriminatory as only one genotype is assigned to serotype 3 and serotype 7F, whereas MLVA4 provides increased discrimination by identifying a number of genetically related but different genotypes.

doi:10.1371/journal.pone.0121870.g004

example ST199 (serotype 19A/15C), ST1012 (serotype 22F/33F) and ST4237 (serotype 6A/6C), however there is no support for capsule switching of the other strains in the international database. This could mean that the capsule switch is relatively new. Alternatively, MLVA4 is too discriminatory and even though true capsule switch occurs, MLVA4 identifies two distinct genotypes therefore the assumption is that no capsule switching has occurred. Further investigation is required to determine whether MLVA4 could fail to detect capsule switches or that MLST is detecting false capsule switches. Since MLVA4 is highly discriminatory, it may enable detection of capsule switching earlier than MLST would. Accurately and quickly detecting relationships between serotypes may have an impact on the selection of serotypes for future vaccine strategies. Furthermore, the ability to examine CC with higher discrimination using MLVA4 can provide insight into which BOX elements are changing. Serotype 7F (MLVA CC1 or ST191) largely diversifies due to BOX-10 and serotype 3 diversifies due to Spneu17. It is unknown what the specific functions of these elements are, however it is known that VNTRs and BOX elements play a role in bacterial competence and virulence and can influence gene expression downstream [30, 43]. VNTR loci with high diversity (e.g. Spneu17) would allow increased discrimination within localised or short-term studies, whereas VNTR loci with low diversity (e.g. BOX-02 and BOX-11) would allow identification of long term changes.

In conclusion, we have developed a MLVA4 method for genotyping invasive *S. pneumoniae*. The main advantage of this new method over other MLVA protocols is the ability to achieve complete MLVA profiles for serotypes while also maintaining a highly discriminatory and fast genotyping technique. Loci that failed to amplify were found to be serotype specific, which may indicate that these BOX elements in these serotypes may be variable and have the capacity to transpose. Further research is required to understand the VNTR genetics of these serotypes as VNTRs and BOX loci are thought to play a role in virulence. MLVA4 is also more suitable for genotyping *S. pneumoniae* than MLST as a more diverse population can be visualised and allow accurate tracking of strains across the state. MLST may be more suitable for a national study, rather than state. This study, establishes a population structure prior and post 13vPCV introduction in Queensland, and it is expected that future monitoring will comprehensively and accurately depict the changes in the pneumococcal population. The future perspective of MLVA is that it will emerge as a cheap and fast genotyping method for localised and national studies that can still be used in conjunction with the currently traditional and slower serotyping and MLST methods for characterising *S. pneumoniae*.

Supporting Information

S1 Fig. Gel electrophoresis of serotype 33F *S. pneumoniae* strains. The white box highlights a large >2000bp BOX-13 fragment in this serotype, identified for isolate number 105, 217, 219, 302, 303, 345, 365 and 367. Isolates 15, 150, 218 and 366 contain fragment lengths of 450bp for BOX-13. A negative control (-ve) is included and size ladders are in beginning and end lanes (100bp ladder and 123bp ladder, respectively.
(TIF)

S1 Table. Genotyping data results for MLST, MLVA1, MLVA2 and MLVA4 applied to Queensland invasive *S. pneumoniae* isolates from 2007 to 2012.
(XLSX)

Acknowledgments

We acknowledge the assistance of Lawrence Ariotti, Ricki Graham and Megan Staples from the Queensland Health Pneumococcal Reference Laboratory, Queensland Health Forensic and Scientific Services. We acknowledge the use of the pneumococcal MLST database which is located at Imperial College London and is funded by the Wellcome Trust.

Author Contributions

Conceived and designed the experiments: RER FH. Performed the experiments: RER. Analyzed the data: RER JS LH FH. Contributed reagents/materials/analysis tools: RER JS FH. Wrote the paper: RER JS LH FH.

References

1. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet 2009; 374: 893–902. doi: 10.1016/S0140-6736(09)61204-6 PMID: 19748398
2. NNDSS (2013). Disease notification rates, QLD, 1991 to 2012 and year-to-date notifications for 2013. Australian Government. Available: http://www9.health.gov.au/cda/source/cda-index.cfm. Accessed 2013 August, 12
3. Lehmann D, Willis J, Moore HC, Giele C, Murphy D, Hoeft A, et al. The changing epidemiology of invasive pneumococcal disease in Aboriginal and non-Aboriginal Western Australians from 1997 through 2007 and emergence of nonvaccine serotypes. Clin Infect Dis 2010; 50: 1477–1486. doi: 10.1086/652440 PMID: 20420501

4. Centers for Disease Control Prevention (2013). Invasive pneumococcal disease (IPD) (*Streptococcus pneumoniae*, invasive disease) (*Streptococcus pneumoniae*): 2010 Case definition. Available: http://wwwn.cdc.gov/NNDSS/script/casedef.aspx?CondYrID=736&DatePub=1/1/2010%2012:00:00%20AM. Accessed 2013 October 23.
5. Spratt BG, Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? Lancet 2000; 356: 1210–1211. PMID: 11072934
6. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. 2007; 3: e168. PMID: 18020702
7. Wyres KL, Klugman KP, Parkhill J, Hakenbeck R, Bentley SD, Brueggemann AB, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis 2013; 207: 439–449. doi: 10.1093/infdis/jis703 PMID: 23175765
8. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genetics 2006; 2: e31. PMID: 16532061
9. Ko KS, Baek JY, Song JH. Capsular Gene Sequences and Genotypes of "Serotype 6E" *Streptococcus pneumoniae* Isolates. Journal of Clinical Microbiology 2013; 51: 3395–3399. doi: 10.1128/JCM.01645-13 PMID: 23824778
10. Calix JJ, Nahm MH (2010) A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. The Journal of Infectious Diseases 202: 29–38. doi: 10.1086/653123 PMID: 20507232
11. Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, et al. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. Journal of Biological Chemistry 2012; 287: 27885. doi: 10.1074/jbc.M112.380451 PMID: 22736767
12. Bratcher PE, Kim K, Kang JH, Hong JY, Nahm MH. Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. Microbiology 2010; 156: 555–560. doi: 10.1099/mic.0.034116-0 PMID: 19942663
13. Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MCC, Nahm MH. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. Journal of Clinical Microbiology 2007; 45: 1225–1233. PMID: 17267625
14. Ansaldi F, Canepa P, de Florentiis D, Bandettini R, Durando P, Icardi G. Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine. Clin Vaccine Immunol 2011; 18: 343–345. doi: 10.1128/CVI.00383-10 PMID: 21177921
15. Kelly T, Dillard JP, Yother J. Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. Infect Immun 1994; 62: 1813–1819. PMID: 8168944
16. Sabharwal V, Stevenson A, Figueira M, Orthopoulos G, Trzciński K, Pelton SI. Capsular switching as a strategy to increase pneumococcal virulence in experimental otitis media model. Microbes Infect, 2013; in press.
17. Selvey C. Information on the Prevenar 13 Supplementary Dose Program. In: Health Q, editor. Brisbane. 2011;
18. Statens Serum Institut (2013). Pneumococcus antisera. Available: http://www.ssi.dk/English/SSI%20Diagnostica/Products%20from%20SSI%20Diagnostica/Antisera_antibodies/Pneumococcus%20antisera.aspx. Accessed 2014 March 20.
19. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiol 1998; 144: 3049–3060.
20. Elberse KEM, Nunes S, Sa-Leao R, van der Heide HGJ, Schouls LM. Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. PLoS One. 2011; 6: e19668. Available at: <Go to ISI>://WOS:000291052200011. Accessed 2012 May 21. doi: 10.1371/journal.pone.0019668 PMID: 21637335
21. Koeck J, Nanjanpop-Lafourcade B, Cade S, Varon E, Sangare L, Valjevac S, et al. Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. BMC Microbiol 2005; 5: 66–74. PMID: 16287512
22. Van Cuyck H, Pichon B, Leroy P, Granger-Farbos A, Underwood A, Soullie B, et al. Multiple-Locus Variable-Number Tandem-Repeat Analysis of *Streptococcus pneumoniae* and comparison with Multiple Loci Sequence Typing. BMC Microbiol 2012; 12: 241–258. doi: 10.1186/1471-2180-12-241 PMID: 23088225
23. Rakov AV, Ubukata K, Ashley Robinson D. Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus *boxB* sequence typing. Infect Genet Evol 2011; 11: 1929–1939. doi: 10.1016/j.meegid.2011.08.016 PMID: 21888992

24. Lindstedt B, Torpdahl M, Vergnaud G, Le Hello S, Weill F, Tietze E, et al. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill 2013; 18(4): pii = 20385. PMID: 23369395
25. Schouls L, van der Heide H (2012). Streptococcus pneumoniae. National Institute for Public Health and the Environment: Ministry of Health, Welfare and Sport. Available: http://www.mlva.net/spneumoniae/default.asp. Accessed 12 April 2012.
26. Hermans PW, Sluijter M, Hoogenboezem T, Heersma H, van Belkum A, de Groot R. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. J Clin Microbiol 1995; 33: 1606–1612. PMID: 7650196
27. Malachowa N, Sabat A, Gniadkowski M, Krzyszton-Russjan J, Empel J, Miedzobrodzki J, et al. Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, *spa* typing, and multilocus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. J Clin Microbiol 2005; 43: 3095–3100. PMID: 16000419
28. Lindstedt B. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis 2005; 26: 2567–2582. PMID: 15937984
29. Yaro S, Lourd M, Traore Y, Njanpop-Lafourcade BM, Sawadogo A, Sangare L, et al. Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. Clin Infect Dis 2006; 43: 693–700. PMID: 16912941
30. Martin B, Humbert O, Camara M, Guenzi E, Walker J, Mitchell T, et al. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res 1992; 20: 3479–3483. PMID: 1630918
31. Croucher NJ, Vernikos GS, Parkhill J, Bentley SD. Identification, variation and transcription of pneumococcal repeat sequences. BMC Genomics 2011; 12: 120–120. doi: 10.1186/1471-2164-12-120 PMID: 21333003
32. Pichon B, Moyce L, Sheppard C, Slack M, Turbitt D, Pebody R, et al. Molecular typing of pneumococci for investigation of linked cases of invasive pneumococcal disease. J Clin Microbiol 2010; 48: 1926–1928. doi: 10.1128/JCM.02054-09 PMID: 20164267
33. Jefferies J, Clarke SC, Diggle MA, Smith A, Dowson C, Mitchell T. Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. Mol Biotechnol 2003; 24: 303–307. PMID: 12777696
34. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrico JA. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 2012; 13: 87. doi: 10.1186/1471-2105-13-87 PMID: 22568821
35. Severiano A, Pinto FR, Ramirez M, Carriço JA. Adjusted Wallace coefficient as a measure of congruence between typing methods. J Clin Microbiol 2011; 49: 3997–4000. doi: 10.1128/JCM.00624-11 PMID: 21918028
36. Wallace DL. A Method for Comparing Two Hierarchical Clusterings: Comment. Journal of the American Statistical Association 1983; 78: 569–576.
37. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 1988; 26: 2465–2466. PMID: 3069867
38. Mehr S, Wood N. Streptococcus pneumoniae—a review of carriage, infection, serotype replacement and vaccination. Paediatric respiratory reviews 2012; 13: 258–264. doi: 10.1016/j.prrv.2011.12.001 PMID: 23069126
39. Johnson DR, D'Onise K, Holland RA, Raupach JCA, Koehler AP. Pneumococcal disease in South Australia: vaccine success but no time for complacency. Vaccine 2012; 30: 2206–2211. doi: 10.1016/j.vaccine.2011.12.119 PMID: 22273663
40. Elberse KEM, van de Pol I, Witteveen S, van der Heide HGJ, Schot CS, van Dijk A, et al. Population structure of invasive *Streptococcus pneumoniae* in The Netherlands in the pre-vaccination era assessed by MLVA and capsular sequence typing. 2011;PLoS One. 6: e20390. doi: 10.1371/journal.pone.0020390 PMID: 21637810
41. Sanchez-Beato AR, Lopez R, Garcia JL. Molecular characterization of PcpA: a novel choline-binding protein of *Streptococcus pneumoniae*. FEMS Microbiol Lett 1998; 164: 207–214. PMID: 9675866
42. Hoang V-T, Gaëlle C, Katia H, Brigitte F, Harriet C, Annick C, et al. Multiple-Locus Variable-Number Tandem-Repeat Analysis for Longitudinal Survey of Sources of *Pseudomonas aeruginosa* Infection in Cystic Fibrosis Patients. J Clin Microbiol 2007; 45: 3175–3183. PMID: 17699654
43. Knutsen E, Johnsborg O, Quentin Y, Claverys J-P, Havarstein LS. BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. J Bacteriol 2006; 188: 8307–8312. PMID: 16997972

Appendix B11: Complete draft manuscript "Changing population structure of invasive *Streptococcus pneumoniae* in Queensland children pre- and post-13vPCV", prepared for PLoS One

1 Changing population structure of invasive *Streptococcus pneumoniae* in
2 Queensland children pre- and post-13vPCV
3
4 Rachael E. Rayner[1,2], John Savill[2], Louise Hafner[1], Flavia Huygens[1]
5 [1]Biomedical Sciences, Queensland University of Technology, Brisbane,
6 Australia.
7 [2]Public Health Microbiology Laboratory, Queensland Health Forensic and
8 Scientific Services, Brisbane, Australia.
9 Corresponding author: Flavia Huygens, IHBI, QUT 60 Musk Ave Kelvin Grove
10 4059, (+617) 3138 0453, f.huygens@qut.edu.au
11 Alternate corresponding author: Rachael E. Rayner, IHBI, 60 Musk Ave Kelvin
12 Grove 4059, (+617) 3138 6310,
13 rachael.rayner@student.qut.edu.au
14
15
16
17
18
19
20
21
22
23
24
25

Abstract

Aims: Changes in the global *Streptococcus pneumoniae* population structure have been observed since the introduction of a childhood vaccine, the 7-valent pneumococcal conjugate vaccine (7vPCV). This has been driven by serotype replacement and capsule switching, particularly in first world countries where vaccines and antibiotics are widely used. A new childhood vaccine, 13-valent pneumococcal conjugate vaccine (13vPCV), was recently introduced in Australia in July 2011, targeting an extra six serotypes not previously targeted. The aim of this study was to determine whether the pneumococcal population structure is changing since the introduction of the 13vPCV in Queensland.

Materials and Methods: A snapshot of the Queensland pneumococcal population in children under 15 years from 2007-2012 was created. Isolates of invasive *S. pneumoniae* submitted to the Pneumococcal Reference Laboratory at Queensland Health Forensic and Scientific Services (QHFSS), Brisbane, were genotyped using a modified Multi-Locus Variable Number of Tandem Repeat Analysis (MLVA) method we developed recently.

Results: Between the periods 2007-2009 and 2010-2012, the 7vPCV serotypes have declined to 6% from 23%, however 13vPCV serotypes have increased from 40% to 61%. There were no significant increases of non-13vPCV serotypes. Potential capsule switching has been detected.

Conclusion: Continued monitoring of the changing pneumococcal population is important.

Keywords: *Streptococcus pneumoniae*, MLVA, invasive, epidemiology, Queensland, 13vPCV, capsule switching

Introduction

Streptococcus pneumoniae is a common but potentially deadly bacterium that causes a wide range of diseases including invasive pneumococcal disease (IPD), pneumonia, otitis media, meningitis and bacteraemia [1]. Young children (under 2 years) and the elderly (over 65 years) are more susceptible to pneumococcal infections [2]. To combat increased disease rates in young children, a 7vPCV (Wyeth Pharmaceuticals Inc.) was developed and implemented worldwide for children under two years, including Australia (introduced 2001 and freely available for all children in 2005). The 7vPCV targets seven serotypes (serotypes 4, 6B, 9V, 14, 18C, 19F and 23F) out of 95 known serotypes [3-8]. Since the introduction of this childhood vaccine, a decline in circulating 7vPCV-serotypes and associated IPD has decreased globally, for instance there was a 78% drop in IPD caused by 7vPCV serotypes in American children under two by 2003 after the 7vPCV was implemented in 2000 [9-10]. Similar decreases of reported IPD caused by 7vPCV-serotypes have been observed in Australia [11-15].

However, non-7vPCV serotypes started to increase globally due to two phenomena; serotype replacement and capsule switching. Serotype replacement is defined as the increase in the incidence of invasive disease caused by non-vaccine serotypes after the introduction of a vaccine [16]. Capsule switching is described as the horizontal transfer of capsule genes from one serotype to another so that the expressed capsule changes from a vaccine-serotype to a non-vaccine serotype, thus becoming “immune” to the effects of the vaccine [17]. More alarmingly, capsule switching events have lead to vaccine escape pneumococcal isolates. In USA, a vaccine escape

pneumococcus ST695 or recombinant P1 has emerged after a serotype 4 switched its capsule to a serotype 19A and spread westward across the continent [17]. ST695 has also been detected in Italy [18]. The transformation of a capsule has been known to affect the virulence properties of a pneumococcal strain [19-21]. No vaccine escape isolates have been reported in Australia; however serotype replacement has been reported [22]. IPD notification rates in Queensland have increased from 6.0 per 100 000 to 7.6 per 100 000 from 2009 to 2012 (http://www9.health.gov.au/cda/source/cda-index.cfm). This is possibly due to the increase of non-7vPCV serotypes in Queensland.

To combat changing pneumococcal populations and increasing invasive diseases, the Wyeth pharmaceutical company developed a 13vPCV to replace the 7vPCV. This decision and selection of serotypes was based on international pneumococcal serotype surveillance data [23]. In July 2011 the 13vPCV was added to the National Vaccine Immunisation Schedule in Australia, targeting an extra six serotypes not covered by the earlier 7vPCV [24]. Since the introduction of 13vPCV in Australia, there have been no publications examining the effect of this new vaccine on the pneumococcal population structure in Queensland or Australia. A 15vPCV is currently on trial targeting an extra two serotypes (serotype 22F and 33F), and may be a potential future vaccine. This study aims to determine the population structure of invasive *S. pneumoniae* in Queensland children under 15 from January 2007 to December 2012. The objective of this study is to determine whether the population structure of non-13vPCV serotypes is changing in Queensland.

Materials and Methods

Setting

The state of Queensland, Australia, has a population of 4.6 million people spread across an area of ~1.7 million km^2 (http://www.oesr.qld.gov.au/products/briefs/pop-growth-qld/qld-pop-counter.php). The National Notifiable Diseases Surveillance System (NNDSS) recorded a 7.6 per 100 000 notification rate of confirmed cases of invasive pneumococcal disease in Queensland in 2012. All cases of IPD are reported by general practitioners and public and private hospitals to the State Communicable Disease Department in the Department of Health, where it is collated. The national data is corrected annually. All Queensland invasive pneumococcal isolates are required to be sent to the Public Health Microbiology Laboratory at Queensland Health Forensic and Scientific Services. There can be a difference between the total number of IPD cases reported and the total number of pneumococcal serotypes reported for each interval and year.

Laboratory methods

S. pneumoniae isolates were serotyped using Quellung reaction at the Pneumococcal Reference Laboratory, Queensland Health Forensic and Scientific Services. An invasive pneumococcal isolate by definition is one that is identified as isolated from normally sterile body sites such as blood, cerebrospinal fluid, tissues or pleural fluid (http://wwwn.cdc.gov/NNDSS/script/casedef.aspx?CondYrID=736&DatePub=1/1/2010%2012:00:00%20AM). Invasive *S. pneumoniae* isolated from children

15 years or younger from January 2007 to December 2012 were resuscitated, isolated and re-cultured to ensure colonies were genetically identical. Crude DNA was extracted by boiling pneumococcal colonies in 400µL TE buffer (~pH8.0) for eight minutes. These were stored at -20°C until further use.

Multi-Locus Variable-number of Tandem Repeat Analysis (MLVA) was performed as previously described [25]. Briefly, eight fluorescently labelled BOX MLVA primers were used to amplify invasive *S. pneumoniae* isolates using PCR and Qiagen Mastermix reagent (Qiagen, Australia). An AB3130 sequencer (Applied Biosystems, Australia) was used to size PCR fragment products with a LIZ1200 internal size ladder (Applied Biosystems, Australia).

The modified MLVA4 method (submitted paper) was also applied. Briefly ten fluorescently labelled VNTR primers were used to amplify invasive *S. pneumoniae* isolates using PCR, and Qiagen Mastermix reagent (Qiagen, Australia). An AB3130 sequencer (Applied Biosystems, Australia) was used to size PCR fragment products with a LIZ1200 internal size ladder (Applied Biosystems, Australia).

Multi-Locus Sequence Typing (MLST) was also performed on selected isolates, namely non-13vPCV serotypes. These isolates were selected because important changes in the pneumococcal population structure will be associated with these isolates. MLST was performed as previously described [26], using the Corbett Cas 1200 liquid handling robot (Qiagen, Australia) for automated sample and mastermix additions [27].

151 ***Analysis***

152 MLVA results were analysed on PeakScanner V1.0 software (Applied
153 Biosystems, Australia). A MLVA type (MT) was assigned to each isolate using
154 the MLVA database (http://www.mlva.net). For our MLVA method, a database
155 was devised and MT types were assigned to each isolates from this database.
156 The population structure for MLVA was analysed by cluster analysis software
157 (eBurst) and displayed as a diagram produced from the MLST database
158 (http://www.mlst.net). The eBurst diagram attempts to identify related
159 genotypes in a population and the founding genotype of mutually exclusive
160 groups [28].

161 MLST results were analysed using ChromasPro software
162 (Technelysium Pty Ltd.) for batch alignment analysis. Sequence types (ST)
163 were assigned to each isolate using the MLST database. The MLST
164 population structure was displayed as an eBurst diagram.

165 Simpson's Index of Diversity (D) was also calculated to compare the
166 discrimination ability of MLVA and MLST methods (accessed from
167 http://alyoung.com/labs/biodiversity_calculator.html).

168

169 **Results**

170 ***Epidemiology of invasive pneumococci***

171 A total of 317 invasive *S. pneumoniae* isolates were genotyped using
172 MLVA4 method. These isolates were sent to QHFSS, Brisbane, and had been
173 collected from children 15 years or younger from January 2007 until
174 December 2012. When examining invasive pneumococcal isolates referred to
175 the laboratory from children 15 years or younger, 68% of isolates where

associated with children under five years. The total number of pneumococcal isolates has declined from 147 (37%) to 125 (31%) from 2007 to 2012 in children under 5 years (Table 1), however isolates have increased in older children from 44 (11%) to 81 (20%) from 2007 to 2012. Upon closer examination, the number of 7vPCV-serotypes observed has been decreasing from 2007 through to 2011 across all age groups (Figure 1). There were a high number of 13vPCV-serotypes isolated from younger children (0-4 years) compared to older children.

In 2011, a sudden increase of 13vPCV-serotypes was observed (65 isolates in 2011 compared to 33 isolates in 2010), however no single serotype was associated with this increase (Figure 1). Interestingly, in the 0-4 year age group, there is a sudden decrease of 13vPCV serotypes in 2012 (n=17) compared to earlier years (average 30 isolates per year). In contrast, the number of 13vPCV-serotypes has increased in older children in 2011-2012.

There are no significant increases in the number of 15vPCV serotypes (serotypes 22F and 33F) or non-15vPCV serotypes. However 30% of isolates in 2010-2012 were caused by non-vaccine serotypes.

Pneumococcal population structure

The overall MLVA eBursts (MLVA1 and MLVA4) for 317 pneumococcal isolates demonstrates which invasive *S. pneumoniae* isolates have been circulating throughout Queensland children from 2007 to 2012 (Figures 2 and 3). The most common pneumococcal serotypes in this study were serotypes 1 (9%), 3 (5%), 6C (3%), 7F (9%), 15C (3%), 19A (36%), 19F (4%) and 33F (3%) (Table 1). Nine percent of these isolates (serotype 6C, 15C and 33F) are not targeted by the 13vPCV. Serotype 19F remains in circulation in

Queensland despite being targeted by the 7vPCV which was freely available in 2005. These 19F serotypes are all multi-drug resistant (unpublished data). Overall serotype replacement has not been observed.

A number of MLVA4 genotypes were observed to be more prevalent than other genotypes of the same serotype. These include MT36 (serotype 1; n= 15), MT35 (serotype 1; n= 10), MT60 (serotype 19A; n = 10), MT59 (serotype 19A/15C; n= 8), MT27 (serotype 7F); n = 7), MT22 (serotype 7F; n= 6), MT21 (serotype7F; n= 4), and MT107 (serotype 6C; n= 4).

The discrimination ability of MLVA1 method, MLVA4 method and MLST were compared. Differences were observed between these three methods. The MLVA4 method had the highest Simpson's Index of Diversity (D = 0.98), whereas MLVA1 method was similar (D = 0.97) and MLST was the lowest (D = 0.87). The population structures by the three methods were also compared. Using MLVA1 method (Figure 2), there are a total of 22 clonal clusters (CC) of which there are three noticeably large clusters; CC1 (predominantly serotypes 7F, 3 and 1), CC2 (serotypes 19A, 15B, 15C and 22F), and CC3 (serotypes 19A and 18A). CC1 is of particular interest as it appears to contain three sub-clusters each with a founding genotype: MT473 (serotype 7F), MT251 (serotype 3) and MT261 (serotypes 1 and 4), however when examining the eBurst diagrams of MLVA4 method (Figure 3), the sub-clusters disappear, and form three separate clusters; CC1 (serotype 7F), CC3 (serotype 3) and CC8 (serotype 1). The MLVA4 method detects a total of 32 clonal clusters. This indicated that MLVA4 method is more discriminatory than MLVA1 method.

The eBurst diagrams also allow examination of the changes in the population structure over time. The MLVA4 method shows that from 2007 to 2009, there are 17 clonal clusters identified and 40% of isolates were 13vPCV serotypes while 23% were 7vPCV serotypes (Figure 4). The 7vPCV was offered freely on the National Vaccine Immunisation Schedule from January 2005 (available from 2001 but was expensive) and the 13vPCV from July 2011. When comparing 2007-2009 data to 2010-2012 we can see that the population structure has changed although there are still 17 clonal clusters (Figure 5). The 7vPCV-serotypes have significantly declined to 6% of the population however 13vPCV-serotypes have increased to 61% of the population. The 13vPCV and 15vPCV target all of the serotypes in 11 clusters (CC1, CC2, CC3, CC5, CC8, CC12, CC13, CC16, CC17, CC20, CC22, and CC29). They do not target all of the serotypes observed in 20 clusters using MLVA4 (CC4, CC6, CC9, CC10, CC14, CC18, CC21, CC23, CC25, CC27, and CC30).

eBurst analysis of MLVA4 data indicates that seven predominant clonal clusters were present during 2007-2012 (Figure 4 and 5), and these include CC1 (serotype 7F), CC2 (serotype 19A and 19F), CC3 (serotype 3), CC5 (serotype 19A), CC6 (serotypes 19A, 15B and 15C), CC10 (serotype 23A) and CC12 (serotype 33F). Some of these clusters are highly clonal, containing a single serotype and few MLVA genotypes such as CC1 (serotype 7F) and CC5 (serotype 19A), whereas the other clusters contain more than one serotype with diverse MLVA genotypes, such as CC6 (serotypes 19A, 15B and 15C), CC22 (serotypes 22F and 33F) and CC2 (serotypes 19A and 19F).

There is also evidence of emerging and disappearing clusters in the pneumococcal population structure. The disappearing clusters that were present in 2007-2009 but not observed in 2010-2012 include CC4 (serotype 38), CC7 (serotype 18B and 18C), CC13 (serotype 19A), CC15 (serotype 23F), CC16 (serotype 1), CC21 (serotype 8), CC22 (serotype 22F), CC23 (serotype 10A), CC26 (serotype 23F) and CC30 (serotype 10F). Some of these clusters contain serotypes targeted by the 7vPCV (CC7, CC15 and CC26). However, other clusters have disappeared that were not targeted by any vaccines including CC4, CC21 and CC30.

On the other hand a number of clusters have emerged in 2010-2012 including CC8 (serotypes 1 and 4), CC9 (serotypes 6C), CC14 (serotype 15C), CC17 (serotype 19A), CC18 (serotype 23B), CC20 (serotype 19A), CC25 (serotype 6C), CC27 (serotype 15C), CC29 (serotype 22F), and CC32 (serotypes 19F). All of these clusters (except for CC8, CC17, CC20 and CC32) contain serotypes that are not targeted by any current vaccine. The number of isolates detected is low, however the number of serotype 6C pneumococci has increased from one isolate (2007-2009) to 11 isolates (2010-1012), and serotype 23B has increased from one isolate (2007-2009) to six isolates (2010-2012).

Examination of the top six clonal clusters and other important clusters

Important clonal clusters observed in MLVA4 data were examined closely for increasing or decreasing clonality. A clonal cluster was defined as containing MLVA genotypes that are single locus variants (SLV) and double locus variants (DLV). Increasing clonality is seen as a tendency towards all

genotypes being single locus variants. Other observations include the number of isolates detected within each study period.

Clonal cluster 1 – This clonal cluster only contains serotype 7F, and with dominant MLVA4 genotypes MT21 (n=4), MT32 (n=3), MT22 (n=6) and MT27 (n=9). Serotype 7F was the second most common serotype observed, below serotype 19A. This cluster has persisted from 2007 through to 2012, decreasing in clonality in 2010-2012 (i.e. becoming more diverse). The assigned founding genotype is MT32.

Clonal cluster 2 – This cluster contains serotype 19A and 19F, both of which are targeted by the 13vPCV. Serotype 19F is also targeted by the original 7vPCV. This cluster has remained in circulation in Queensland children from 2007 through to 2012. The cluster has shifted to mainly contain serotype 19A rather than 19F. This is most likely due to the serotype 19F being targeted by the 7vPCV. Despite the freely available 7vPCV implemented in 2005, serotype 19F remains in circulation. There is no founding genotype in this cluster. MT101 (serotype 19F) has been observed in both 2007-2009 and the 2010-2012 periods.

Clonal cluster 3 – This cluster only contains serotype 3, and has persisted from 2007 through to 2012 without any major changes in the MLVA types in the clonal cluster. Serotype 3 is the fourth highest IPD causing serotype in Queensland children.

Clonal cluster 4 – This cluster only contains serotype 38, which is not targeted by any current vaccine. This cluster was dominant in earlier years (2007 – 2009).

297 ***Clonal cluster 5*** – This cluster only contains serotype 19A, and all
298 isolates are single locus variants of the dominant genotype MT60. This cluster
299 has also persisted from 2007-2012, and has actually decreased in clonality in
300 2010-2012. Of particular interest is the predicted founding genotype, MT60,
301 which makes up most of the referred isolates in CC2 (10 pneumococcal
302 isolates in 2010-2012). CC2 is more genetically related to CC1 (serotype
303 19A/15B/15C) as four MLVA alleles out of 10VNTR are the same. CC2
304 isolates share only two alleles with isolates in CC5 (serotypes 19A and 19F)
305 and CC11 (serotype 19A). CC11 and CC5 are not genetically related as no
306 alleles are common.

307 ***Clonal cluster 6*** - This large cluster contains serotypes 15B, 15C and
308 19A, of which only serotype 19A is targeted by the 13vPCV. CC6 has
309 persisted from 2007 through to 2012. The dominant MLVA4 genotype is MT59
310 (n=11), with a potential capsule switch from a serotype 19A to serotype 15C.
311 The number of isolates has declined in the 2010 to 2012 period, indicating
312 that the cluster is becoming more clonal. Serotype 19A in this cluster is
313 declining in numbers in later years, and the cluster is becoming dominated by
314 serotypes 15B and 15C. Serogroup 15 is not targeted by any childhood
315 vaccine.

316 ***Clonal cluster 8*** – This clonal cluster predominantly contains serotype
317 1 but also contains a potential capsule switch to a serotype 4 (1 isolate). The
318 cluster in 2010-2012 is highly clonal, and the majority of strains in this cluster
319 are MT35 and MT36. CC8 is not clonally related to CC16 (serotype 1)
320 observed in 2007-2009. CC8 appears to have emerged in 2010-2012 either

from a distantly related strain of CC16 or an independent source. Serotype 1 is the third highest serotype referred to the laboratory.

Clonal clusters 9 and 25 – These clusters contain serotype 6A/6C and 6C, respectively. Both clusters have emerged in 2010-2012 and are clonally unrelated to each other. Serotype 6C is not targeted by the 13vPCV. The in-trial 15vPCV does not target serotype 6C.

Clonal cluster 12 – This small cluster contains serotype 33F, a target of the future vaccine, 15vPCV. The MLVA4 genotypes include MT398 (n=1), MT217 (n=3), MT105 (n=2) and MT365 (n=2). CC12 has persisted from 2007 through to 2012, however it doesn't appear to be changing. Serotype 33F has been isolated more commonly from younger children (<5 years) in this study.

Clonal clusters 14 and 27 – Both of these smaller clusters contain serotype 15C and 15B/15C, respectively. MLVA4 types include MT52 (n=1), MT360 (n=1), MT287 (n=1), MT336 (n=1) and MT53 (n=2). Both clusters have emerged in 2010-2012. These clusters are not clonally related to CC6 (15B/15C and 19A). No vaccine presently targets 15B or 15C.

Clonal cluster 18 – This cluster only contains serotype 23B which is not targeted by the 13vPCV or the in-trial 15vPCV. CC18 has emerged in 2010-2012. The total number of serotype 23B observed has increased from one (2007-2009) to six (2010-2012). There is a potential for serotype 23B to become serotype replacement, and it is not targeted by the 13vPCV and 15vPCV.

Clonal clusters 22 and 29 – CC22 contains serotype 22F and 33F (serotype 33F MT366 is a single isolate in 2010-2012), whereas CC29 only contains serotype 22F. Serotype 22F (MT143, n=1) and 33F (MT366, n=1)

346 are single locus variants, despite that serotype 22F was observed in 2007-
347 2009 and the 33F observed in 2010-2012. CC22 is not clonally related to
348 CC29. Both serotype 22F and 33F are targets of the in-trial 15vPCV, but are
349 not targeted by the 13vPCV.

350 ***Capsule switching***

351 We have observed capsule switches occurring in the Queensland
352 pneumococcal population (Table 2). A capsule switch is defined as two
353 pneumococcal isolates with identical genotypes that have different capsule
354 types or serotypes. The MLVA4 method detects two potential capsule
355 switches (Figure 3), including CC8, between a serotype 1 and serotype 4, and
356 CC6, between a serotype 19A and 15C. These capsule switches were also
357 detected using MLVA1 method and MLST, however these other methods also
358 indicate many other potential capsule switching. MLST indicates that there are
359 four additional capsule switches, including serotype 22F to 33F (ST1012),
360 serotype 18B to 18C (ST1255), serotype 15B to 15C (ST199) and serotype 8
361 to 11A (ST53). MLVA1 indicates that there are five additional capsule
362 switches, including serotype 19A to 19F, serotype 8 to 11A, serotype 18B to
363 18C, and two incidents of serotype 15B to 15C.

364
365 **Discussion**

366 This is one of the first published studies that reports a potential decline
367 in thirteen-valent *S. pneumoniae* serotypes associated with invasive disease
368 in children since the introduction of the 13vPCV. This is a promising result as
369 there was a reported 168% increase in non-7vPCV serotypes in children
370 under five in Australia in 2008 compared to 2002, with a four-fold increase due
371 to serotype 19A [29]. We have also observed a steady decline in 7vPCV-

serotypes causing disease which is as expected. A year after the introduction of the 13vPCV (July 2011), we have observed a decline in the number of 13vPCV pneumococcal isolates from children 0-4 years from 39 isolates to 17 isolates. This may indicate that the new childhood vaccine is already starting to successfully reduce the incidence of IPD in young children, as is being observed elsewhere in the world. In the USA, since the introduction of 13vPCV in 2010, there has been a 53% decline in IPD in children under 24 months compared to the calculated average in 2007-2009. This decline was due to falls in 13vPCV-serotype 19A (58%), serotype 7F (54%) and serotype 3 (68%) [30]. A decline in 13vPCV serotypes was also observed in an Alaska in 2009-2011 [31]. Similarly Picazo *et al.* [32] identified a decrease in IPD caused by 13vPCV serotypes in 2010-2011 compared to 2007-2010 in Madrid, Spain, and a study in German children has shown an overall decrease in 13vPCV-serotypes causing IPD despite an increase of serotype 19A in 2010 and 2011 [33]. A study in French children has observed a significant lowering of carriage rates of serotype 19A, 7F and 6C with a single dose of 13vPCV, compared to children only vaccinated with 7vPCV [34], however lowering carriage rates does not necessarily reflect a decrease in invasive disease.

On the other hand this study has not observed a decline in 13vPCV serotypes in older children; in fact pneumococcal isolates increased in 2011 and 2012. This may be because children older than two years do not have access to the free Government funded vaccination schedule in Australia and are hence susceptible to the 13vPCV serotypes. It is hopeful though that as the 13vPCV is continued to be implemented for children under two, IPD cases

397 will fall in the older age groups as vaccinated children become older. The
398 effects of herd immunity may also decrease IPD cases in older children. Herd
399 immunity is an important factor in reducing IPD cases in various populations
400 [9].

401 In this study, a number of clonal clusters have emerged and
402 disappeared in the Queensland pneumococcal population. Some clonal
403 clusters that have disappeared include serotypes that were not targeted by
404 any pneumococcal vaccine, including CC4 (serotype 38), CC21 (serotype 8)
405 and CC30 (serotype 10F), indicating that the population structure can change
406 due to natural fluctuations and are not necessarily eliminated by vaccines.
407 Due to some clonal clusters being highly clonal and only consisting of a single
408 serotype, it is expected that CC1 (serotype 7F), CC3 (serotype 3) and CC5
409 (serotype 19A) will gradually decline and disappear as they are targeted by
410 the 13vPCV. Since serotype 1 is more commonly observed in older children,
411 the effects of the 13vPCV may not be observed in the short term. Conversely,
412 a number of clonal clusters have emerged, indicating the possible beginnings
413 of serotype replacement, especially those clusters containing serotypes not
414 targeted by the 13vPCV. It will therefore be important to monitor the changes
415 of CC9 (serotypes 6C), CC14 (serotype 15C), CC18 (serotype 23B), CC25
416 (serotype 6C), CC27 (serotype 15C) and CC29 (serotype 22F) for potential
417 changes. Continuing surveillance for changes to these clonal clusters will
418 assist in vaccine health policy decision making and the impact of introducing
419 the 15vPCV (targets serotype 22F and 33F).

420 A number of clonal clusters were observed to contain the same
421 serotype, and yet were not identified to be genetically related. These included

CC5 (serotype 19A), CC6 (serotype 19A/15B/15C), CC13 (serotype 19A), CC17 (serotype 19A) and CC20 (serotype 19A), indicating that these four clusters may have originated from independent sources, or that the clusters have diverged from an original cluster decades ago. It may be that the success of serotype 19A to cause increasing IPD cases after the introduction of the 7vPCV globally could rely on the fact that all strains are not clonally related, as demonstrated by the fact that at least four independent clonal clusters have been identified in Queensland. Since serotype 19A is now targeted by the 13vPCV, it is expected that there will be a decline in IPD cases associated with serotype 19A, except for those isolates that are multi-drug resistant. This could emerge as a similar issue for serotype 22F as we have identified two clonal clusters (CC15 and CC22) that contain genetically unrelated isolates of serotype 22F. Since the 13vPCV does not target 22F, it is important that these clusters are monitored.

Interestingly, despite the introduction of 13vPCV, a number of 7vPCV-serotypes were still being observed; particularly serotype 19F which was the fifth highest most common serotype in this study. The 7vPCV was licensed in 2001 and made freely available for children in 2005, which means most children should have been vaccinated. There is 95% vaccine coverage in non-indigenous citizens and 85% coverage in indigenous citizens across Australia [29]. Due to restricted access to patient data for this study, the vaccination history of 7vPCV is unknown for this study, therefore we are unable to determine whether IPD caused by serotype 19F is due to non-vaccination or a vaccine failure or multi-drug resistance. Another possibility is that certain strains of 19F have developed the ability to evade the vaccine as it has been

447 noted that strain MT101 (19F) has been detected in both 2007-2009 and
448 2010-2012.

449 Overall IPD notification rates will not decline further with continued use
450 of 7vPCV, therefore the introduction of the 13vPCV should provide further
451 decrease of IPD since it targets an extra six serotypes not targeted by the
452 7vPCV. We have already noted an increase of 13vPCV-serotypes from 40%
453 (2007-2009) to 61% (2010-2012) in children under 15 years, indicating the
454 effects of serotype replacement after the 7vPCV was freely available in 2005.
455 In South Australia, it has been reported that increases of non-7vPCV
456 serotypes, with serotype 19A as the dominant replacement serotype, has
457 occurred [35]. These authors also commented that non-7vPCV serotypes may
458 be associated with more severe diseases, as they noticed higher rates of
459 complications in empyema [35]. Oftedah *et al.* [13] reported increases of non-
460 7vPCV serotypes in New South Wales children before (5%) and after (18%)
461 the 7vPCV was introduced.

462 An unusual spike of 13vPCV-serotypes was also detected in 2011. A
463 serotype 1 outbreak occurred in Northern Australia and Queensland, which
464 may have contributed to the unusual spike in IPD cases [36]. Even by omitting
465 serotype 1 isolates that may be a part of an outbreak, the number of 13vPCV
466 isolates are still unusually high. It is not known why the number of 13vPCV-
467 serotypes has significantly increased in 2011, but it appears that the
468 introduction of the 13vPCV has come at an appropriate time. High incidences
469 of IPD in young children have also been reported in other publications [11, 37-
470 38].

Serotype replacement has not been observed since the introduction of the 13vPCV in 2011 in this study, however there have been increases of serotype 6C and 23B, both not targeted by the 13vPCV. Altogether, serotype 6C, 15C and 33F (9% of isolates) could increase by serotype replacement as they are not targeted by the 13vPCV. It is recommended that surveillance by way of MLVA continue to enable reporting of future serotype replacement. Cooper *et al.* [39] has demonstrated that the 13vPCV may have cross-protection against 6C as there was a 96% OPA (opsonaphagocytic assay) response. In South Australia there have also been reports of increases in non-13vPCV serotypes [14].

By comparing the MLVA4 method (unpublished data) with a previously published MLVA method [25] and the "gold standard" MLST, we were able to determine the limitations of using a less discriminatory genotyping method. This is particularly important in determining changes in the population structure and emerging genotypes and capsule switches. The MLVA4 method, having the highest discrimination (D=0.98) compared to MLST (D=0.87), has allowed us to display a more accurate population structure.

Higher discriminatory power can impact interpretations of population structures, especially when identifying capsule switches. A capsule switch is identified when two pneumococcal isolates share the same genetic profile (i.e. same MLVA profile) but have two different serotypes, such as serotype 19A and serotype 4. Our MLVA method identifies less capsule switches than the other two genotyping methods. Other potential capsule switches observed using MLVA1 or MLST may be false positives due to the lower discriminatory power of these genotyping methods. If one wishes to study pneumococcal

496 capsular switches, then using MLVA4 method ensures there are less false
497 positives. Therefore expenses can be minimised and time can be reduced
498 when sequencing or genotyping fewer pneumococcal capsule gene cassettes.
499 The MLVA4 method appears to discriminate most of the clonal clusters
500 as a single serotype, however it should not be applied to serotype by serotype
501 (as other researchers have done), because clonal relationships between
502 different serotypes may be missed, for example in this study, serotypes 11A
503 and 8, 18B and 18C, 22F and 33F, 19A and 19F, and 15B, 15C and 19A all
504 form individual mixed clusters. Potential capsule switching would also not be
505 identified, for example serotype 1 and 4, and serotype 15C and 19A detected
506 in this study. It is important to detect these relationships since serotypes 11A,
507 8, 18B, 22F, 33F, 6C, 15B and 15C are not targeted by the current childhood
508 vaccines. As they are shown to be genetically similar to other serotypes, it
509 may mean they have the ability to switch capsules more easily or become
510 serotype replacement strains.
511
512 **Conclusions and future perspective**
513 Despite the 13vPCV being introduced in July 2011, already results
514 seem to give preliminary indications that this new vaccine is already
515 decreasing 13vPCV serotypes in young children in 2012. This is a positive
516 outcome of the 13vPCV. Unfortunately, an increase of 13vPCV serotypes in
517 the older children was also observed since the introduction of the 13vPCV.
518 Since these children do not have access to the Government funded 13vPCV,
519 it is understandable that 13vPCV-serotypes continue to cause IPD in older
520 children. Hopefully as 13vPCV is continued to be implemented, IPD cases will

drop in this age group as vaccinated children become older. Also the effects of herd immunity may start to take hold, decreasing the likelihood for older siblings or children to contract IPD caused by 13vPCV-serotypes.

The 2007-2012 MLVA genotype data serves as a baseline for future comparisons. It will be important to continue monitoring the population structure for increases in number of referred isolates within CC's and for expansion of clonal complexes.

Executive summary

Objectives

- To determine whether the pneumococcal population structure is changing after the introduction of the 13vPCV in Queensland, Australia.

Methods

- Our previously developed MLVA method (submitted paper) was used to genotype isolates of *S. pneumoniae* originally isolated from Queensland children aged under fifteen.
- MLST was used as a comparison genotyping method.

Results

- Between the periods 2007-2009 and 2010-2012, the 7vPCV serotypes have declined to 6% from 23%.
- The 13vPCV serotypes have increased from 40% to 61%.
- There were no significant increases of non-13vPCV serotypes.
- Potential capsule switching has been detected.

Conclusion

- The 13vPCV seems to be influencing the incidence of invasive pneumococcal diseases in children aged under four.
- Future surveillance of the pneumococcal population is essential since increases in pneumococcal disease has been observed in older children, and the 13vPCV does not protect children against all pneumococcal serotypes that cause disease.

References

1. Koeck J, Nanjanpop-Lafourcade B, Cade S *et al.*: Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. *BMC Microbiology* 5(1), 66-74 (2005). * **First published MLVA method for genotyping *S. pneumoniae*.**
2. Scott JaG, Hall AJ, Dagan R *et al.*: Serogroup-specific epidemiology of *Streptococcus pneumoniae*: associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clinical Infectious Diseases* 22(6), 973-981 (1996).
3. Bentley SD, Aanensen DM, Mavroidi A *et al.*: Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genetics* 2 (3), 10.1371/journal.pgen.0020031:10.1371/journal.pgen.0020031 (2006).
4. Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MCC, Nahm MH: Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology* 45(4), 1225-1233 (2007).

5. Calix JJ, Nahm MH: A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *Journal of Infectious Diseases* 202(1), 29-38 (2010).
6. Calix JJ, Porambo RJ, Brady AM *et al.*: Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *Journal of Biological Chemistry* 287(33), 27885 (2012).
7. Bratcher PE, Kim K, Kang JH, Hong JY, Nahm MH: Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. *Microbiology* 156(2), 555-560 (2010).
8. Ko KS, Baek JY, Song JH: Capsular Gene Sequences and Genotypes of "Serotype 6E" *Streptococcus pneumoniae* Isolates. *Journal of Clinical Microbiology* 51(10), 3395-3399 (2013).
9. Reingold A, Hadler J, Farley MM, Harrison L: Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease - United States, 1998-2003. *JAMA* 294(16), 2022-2026 (2005).
10. Whitney CG, Farley MM, Hadler J *et al.*: Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *The New England Journal of Medicine* 348(18), 1737-1746 (2003).

11. Roche PW, Krause V, Cook H *et al.*: Invasive pneumococcal disease in Australia, 2006. *Communicable Diseases Intelligence* 32(1), 18-30 (2008).
12. Lehmann D, Willis J, Moore HC *et al.*: The changing epidemiology of invasive pneumococcal disease in Aboriginal and non-Aboriginal Western Australians from 1997 through 2007 and emergence of nonvaccine serotypes. *Clinical Infectious Diseases* 50(11), 1477-1486 (2010).
13. Oftadeh S, Gidding HF, Gilbert GL: Laboratory surveillance of invasive pneumococcal disease in New South Wales, Australia, before and after introduction of 7-valent conjugate vaccine: reduced disease, but not antibiotic resistance rates. *Epidemiology and Infection* 141(9), 1797-1806 (2013).
14. Johnson DR, D'onise K, Holland RA, Raupach JCA, Koehler AP: Pneumococcal disease in South Australia: vaccine success but no time for complacency. *Vaccine* 30(12), 2206-2211 (2012).
15. Fitzgerald T, Massey PD, Islam F: Changes in invasive pneumococcal disease serotypes in a regional area of Australia following three years of 7vPCV introduction. *Western Pacific Surveillance and Response Journal* 3(2), 33-38 (2012).
16. Weinberger DM, Malley R, Lipsitch M: Serotype replacement in disease after pneumococcal vaccination. *Lancet* 378(9807), 1962-1973 (2011).
***A review of the phenomenon of serotype replacement since the introduction of the pneumococcal vaccines.**

17. Brueggemann AB, Pai R, Crook DW, Beall B: Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathogens* 3 (11), 10.1371/journal.ppat.0030168:10.1371/journal.ppat.0030168 (2007). ***One of the first descriptions of a vaccine escape pneumococci after switching its capsule.**

18. Ansaldi F, Canepa P, De Florentiis D, Bandettini R, Durando P, Icardi G: Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine. *Clinical and Vaccine Immunology* 18(2), 343-345 (2011).

19. Kelly T, Dillard JP, Yother J: Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infection and Immunity* 62(5), 1813-1819 (1994).

20. Elberse K, Witteveen S, Van Der Heide H *et al.*: Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. *PLoS One* 6 (9), 10.1371/journal.pone.0025018:10.1371/journal.pone.0025018 (2011).

21. Sabharwal V, Stevenson A, Figueira M, Orthopoulos G, Trzciński K, Pelton SI: Capsular switching as a strategy to increase pneumococcal virulence in experimental otitis media model. *Microbes and Infection*, (2013).

22. Hanna JN, Humphreys JL, Murphy DM, Smith HV: Invasive pneumococcal disease in non-Indigenous people in north Queensland, 2001-2009. *Medical Journal of Australia* 193(7), 392-396 (2010).

641 23. Centers for Disease Control Prevention: Licensure of a 13-valent
642 pneumococcal conjugate vaccine (PCV13) and recommendations for
643 use among children - Advisory Committee on Immunization Practices
644 (ACIP), 2010. *Morbidity and Mortality Weekly Report* 59(9), 258-261
645 (2010).
646 24. Selvey C. Information on the Prevenar 13 Supplementary Dose
647 Program. Availed from:
648 http://www.seagp.org.au/documents/Services/PracticeSupport/PublicH
649 ealth/20110602_LETTER_QHIP_to_providers_JS.pdf (Accessed: May,
650 29, 2012).
651 25. Elberse KEM, Nunes S, Sa-Leao R, Van Der Heide HGJ, Schouls LM:
652 Multiple-locus variable number tandem repeat analysis for
653 *Streptococcus pneumoniae*: comparison with PFGE and MLST. *PLoS*
654 *One* 6 (5),
655 10.1371/journal.pone.0019668:10.1371/journal.pone.0019668 (2011).
656 ****A different MLVA protocol used in the Netherlands and adopted**
657 **in this study.**
658 26. Enright MC, Spratt BG: A multilocus sequence typing scheme for
659 *Streptococcus pneumoniae*: identification of clones associated with
660 serious invasive disease. *Microbiology* 144(11), 3049-3060 (1998).
661 ****The original MLST method for genotyping *S. pneumoniae*.**
662 27. Jefferies J, Clarke SC, Diggle MA, Smith A, Dowson C, Mitchell T:
663 Automated pneumococcal MLST using liquid-handling robotics and a
664 capillary DNA sequencer. *Molecular Biotechnology* 24(3), 303-307
665 (2003).

28. Aanensen DM, Spratt BG: The multilocus sequence typing network: mlst.net. *Nucleic Acids Research* 33, 728-733 (2005).
29. Barry C, Krause V, Cook H, Menzies R: Invasive pneumococcal disease in Australia 2007 and 2008. *Communicable Diseases Intelligence* 36(2), (2012).
30. Kaplan SL, Barson WJ, Lin PL *et al.*: Early trends for invasive pneumococcal infections in children after the introduction of the 13-valent pneumococcal conjugate vaccine. *Pediatric Infectious Disease Journal* 32(3), 203-207 (2013).
31. Singleton R, Wenger JD, Klejka JA *et al.*: The 13-valent pneumococcal conjugate vaccine for invasive pneumococcal disease in Alaska native children: results of a clinical trial. *Paediatric Respiratory Reviews* 32(3), 257-263 (2013).
32. Picazo J, Ruiz-Contreras J, Casado-Flores J *et al.*: Impact of introduction of conjugate vaccine in the vaccination schedule on the incidence of pediatric invasive pneumococcal disease requiring hospitalization in Madrid 2007 to 2011. *Pediatric Infectious Disease Journal* 32(6), 656-661 (2013).
33. Van Der Linden M, Reinert RR, Kern WV, Imöhl M: Epidemiology of serotype 19A isolates from invasive pneumococcal disease in German children. *BMC Infectious Diseases* 13(1), 70 (2013).
34. Cohen R, Levy C, Bonnet E *et al.*: Risk factors for serotype 19A carriage after introduction of 7-valent pneumococcal vaccination. *BMC Infectious Diseases* 11(1), 95-95 (2011).

690 35. Tran T, Clarke M, Marshall H: Invasive pneumococcal disease: impact
691 of pneumococcal vaccination in South Australian children. *International*
692 *Journal of Infectious Diseases* 16
693 10.1016/j.ijid.2012.05.1002:10.1016/j.ijid.2012.05.1002 (2012).
694 36. Staples M, Graham R, Ariotti L *et al.*: Molecular characterisation of an
695 Australian serotype 1 *Streptococcus pneumoniae* outbreak. *Australian*
696 *Society of Microbiology*, (2012).
697 37. Slaon-Gardner T, Stirzaker S, Knuckey D *et al.*: Australia's notifiable
698 disease status, 2009: annual report of the National Notifiable Diseases
699 Surveillance System. *Communicable Diseases Intelligence* 35(2), 61-
700 131 (2011).
701 38. Roche P, Krause V, Cook H *et al.*: Invasive pneumococcal disease in
702 Australia, 2005. *Communicable Diseases Intelligence* 31(1), 86-100
703 (2007).
704 39. Cooper D, Yu X, Sidhu M, Nahm MH, Fernsten P, Jansen KU: The 13-
705 valent pneumococcal conjugate vaccine (PCV13) elicits cross-
706 functional opsonophagocytic killing responses in humans to
707 *Streptococcus pneumoniae* serotypes 6C and 7A. *Vaccine* 29(41),
708 7207-7211 (2011).
709

710 **Financial disclosure**

711 This work was supported by the Queensland Health Forensic and Scientific
712 Services, Coopers Plains, Australia.

713

714 **Acknowledgements**

715 We would like to acknowledge the assistance of Lawrence Ariotti, Ricki
716 Graham and Megan Staples from the Queensland Health Pneumococcal
717 Reference Laboratory, Queensland Health Forensic and Scientific Services.

718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736

737 Table 1: Distribution of pneumococcal serotypes causing invasive disease in
738 Queensland children age <5 yr and >5 yrs in the time periods of 2007-2009
739 and 2010-2012.

	2007-2009		2010-2012		
	<5 yr	≥5 yr	<5 yr	≥5 yr	TOTAL
Serotype	N (%)	N (%)	N (%)	N (%)	N (%)
1	1 (0.7)	9 (20)	6 (5)	19 (23)	35 (9)
3	7 (5)	2 (5)	5 (4)	5 (6)	19 (5)
4	1 (0.7)	2 (5)	1 (0.8)	1 (1)	5
6A	3 (2)	1 (2)		1 (1)	5
6B	2 (1)				2
6C		1 (2)	4 (3)	7 (9)	12 (3)
7F	4 (3)	4 (9)	16 (13)	12 (15)	36 (9)
8	4 (3)				4
9N	2 (1)			1 (1)	3
9V	2 (1)			1 (1)	3
10A	1 (0.7)	2 (5)			3
10F	2 (1)				2
11A	5 (4)		1 (0.8)	2 (2)	8
12F		2 (5)	1 (0.8)	1 (1)	4
14	2 (1)	1 (2)		1 (1)	4
15B	2 (1)		3 (2)	1 (1)	6
15C	4 (3)	1 (2)	7 (6)		12 (3)
16F				1 (1)	1
18A	1 (0.7)	1 (2)	1 (0.8)	3 (4)	6
18B	1 (0.7)				1
18C	2 (1)	3 (7)	1 (0.8)		6
19A	68 (48)	5 (11)	57 (46)	15 (19)	145 (36)
19F	8 (6)	1 (2)	5 (4)	1 (1)	15 (4)
22A	1 (0.7)				1
22F	5 (4)		1 (0.8)	1 (1)	7
23A	1 (0.7)	1 (2)	1 (0.8)	2 (2)	5
23B	1 (0.7)		3 (2)	3 (4)	7
23F	3 (2)	4 (9)			7
24F	1 (0.7)				1
33B			1 (0.8)		1
33F	4 (3)	2 (5)	6 (5)		12
34	1 (0.7)			1 (1)	2
35B	2 (1)		2 (2)	1 (1)	5
35F	1 (0.7)		2 (2)	1 (1)	4
38	5 (4)	2 (5)	1 (0.8)		8
TOTAL	147	44	125	81	397

740 Table 2: Potential capsule switches detected in Queensland children by each
741 genotyping method, MLST, Elberse's MLVA and our MLVA methods. The
742 potential capsule switches (boxed) have been consistently detected with all
743 three genotyping methods.

MLST (ST)	Serotype (year observed); number of isolates
1012	22F (2007-2008); n=3 33F (2012);n=1
1255	18B (2007);n=1 18C (2008,2010);n=3
199	15B (2007,2010,2011);n=4 15C (2007,2010,2012);n=4
306	1(2007-2012);n=23 4(2012);n=1
411	19A (2007-2012);n=12 15B (2009);n=1 15C (2007);n=2
4237	6A (2008);n=1 6C (2010-2012);n=6
53	8 (2008-2009);n=3 11A (2011);n=1
Elberse's MLVA (MT)	**Serotype (year observed); number of isolates**
1190	19A (2007-2012);n=3 19F (2007,2010);n=2
1374	6A (2008);n=1 6C (2010-2012);n=5
18	8 (2008-2009);n=3 11A (2011);n=1
2002	18B (2007);n=1 18C (2008-2010);n=3
261	1 (2007-2012);n=15 4 (2012);n=1
58	15B (2011);n=1 15C (2007, 2012);n=3
67	15C (2007);n=1 19A (2007-2012);n=12
904	15B (2012);n=1 15C (2011);n=2
Our MLVA (MT)	**Serotype (year observed); number of isolates**
59	15C (2007) ; n= 1 19A (2007); n = 10
107	6A (2008); n= 1 6C (2010-2012); n= 3
36	1 (2007-2012); n=16 4 (2012); n= 1

744

745

Figure 1: Prevalence of vaccine-specific serotypes isolated from Queensland children from 2007-2012. Pneumococcal serotypes are distinguished as targeted by the 7vPCV, 13vPCV, in-trial 15vPCV or not targeted by any current vaccine (non-15vPCV). The 13vPCV was introduced in July 2011 in Australia for children less than 2 years.

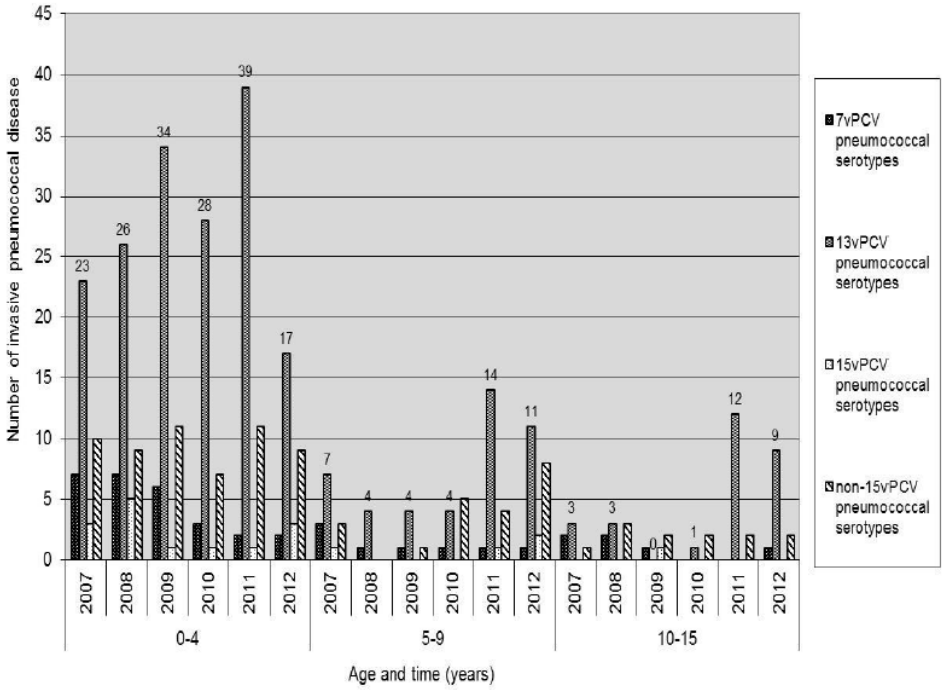
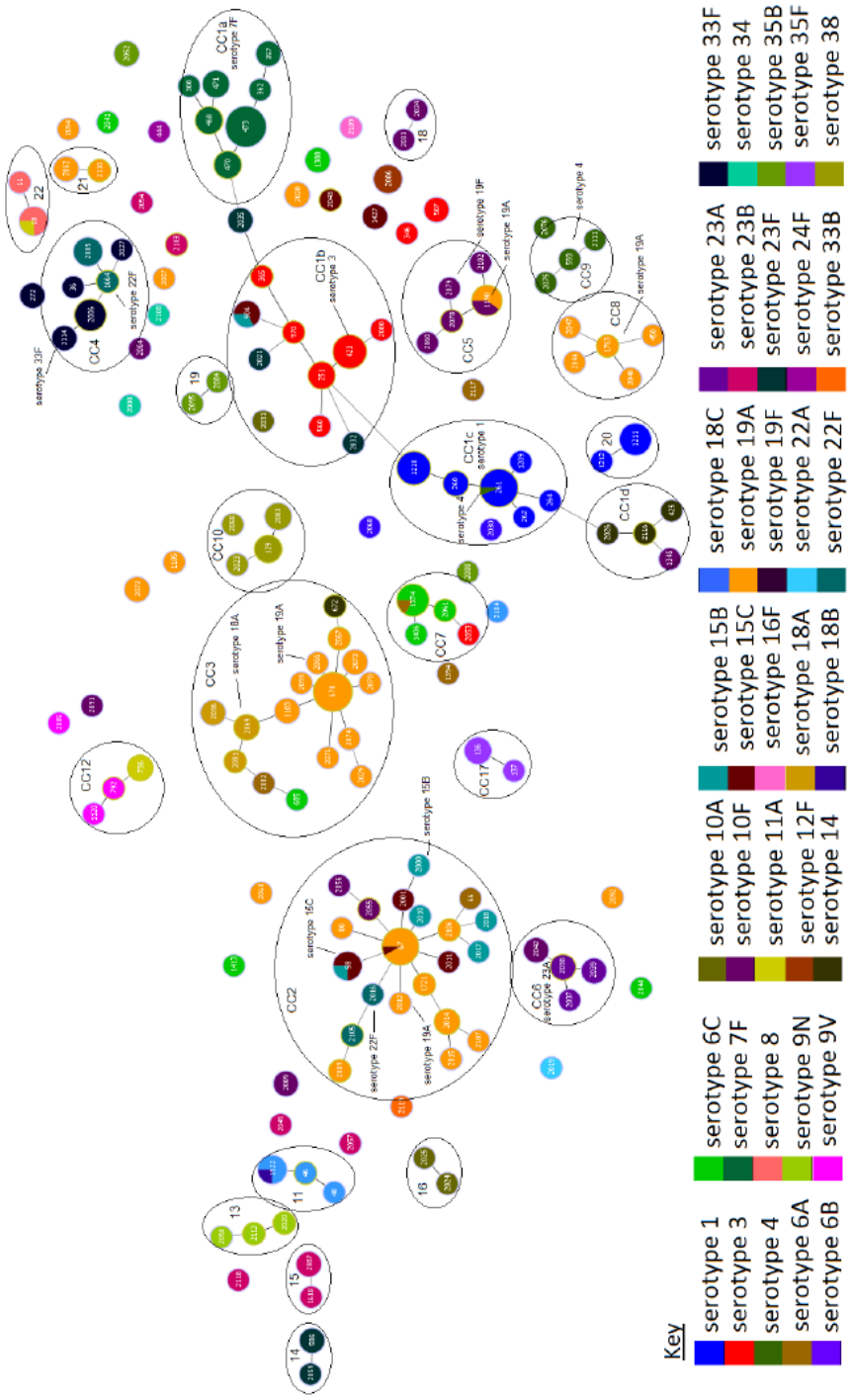


Figure 2: MLVA1 eBurst diagram with serotypes overlayed (n=317). Clonal clusters (CCs) are circled and contain single locus variants (SLV) and double locus variants (DLV). CC1 has been split into four sub-clusters (a, b, c and d). MLVA1 genotype is presented inside each circle, and size of circle represents the number of isolates detected in this study. Each colour represents a different serotype, as shown in the legend key. Circles with split colours

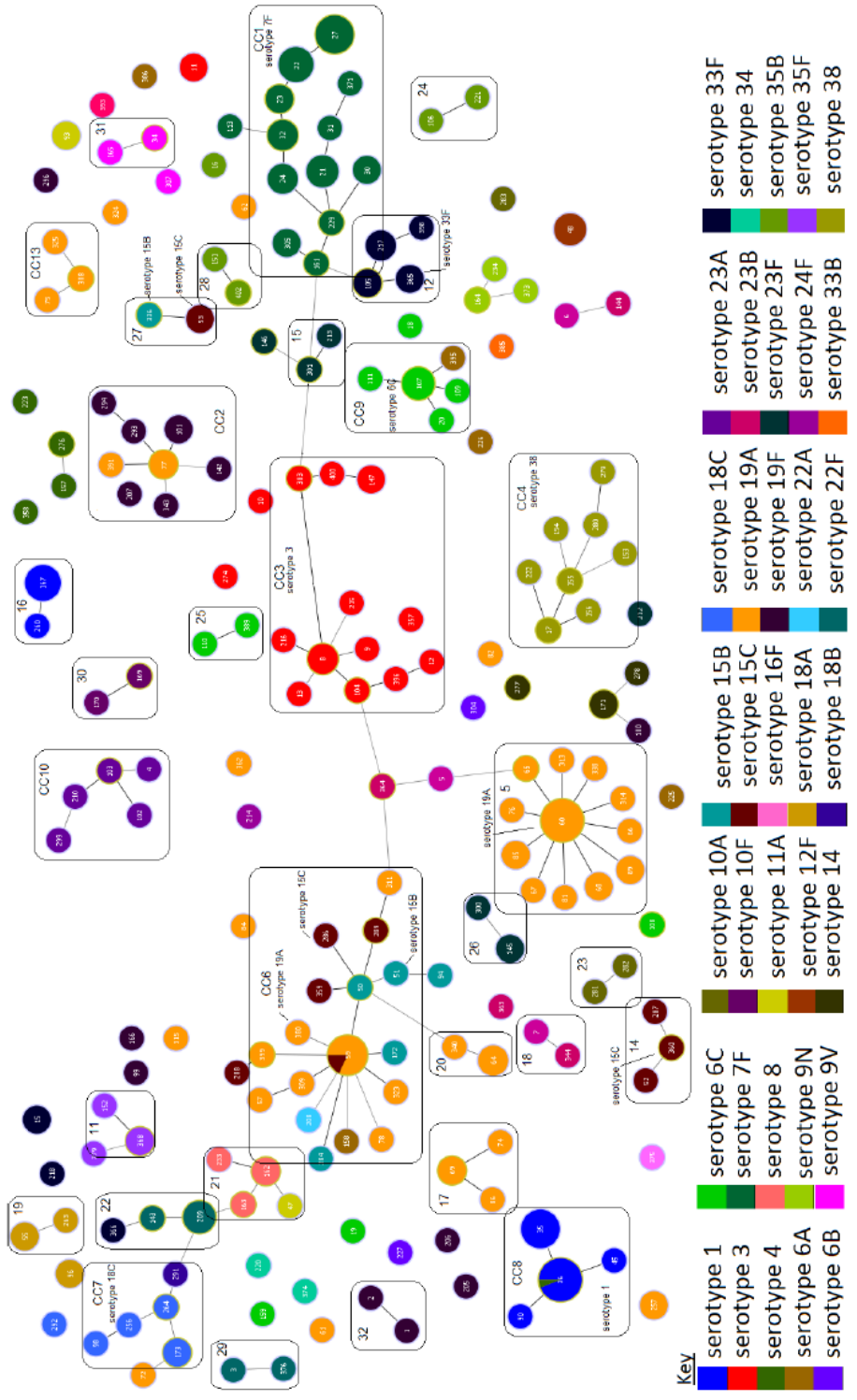
758 indicate that two or more serotypes have the same MLVA1 genotype,

759 indicating potential capsule switching.



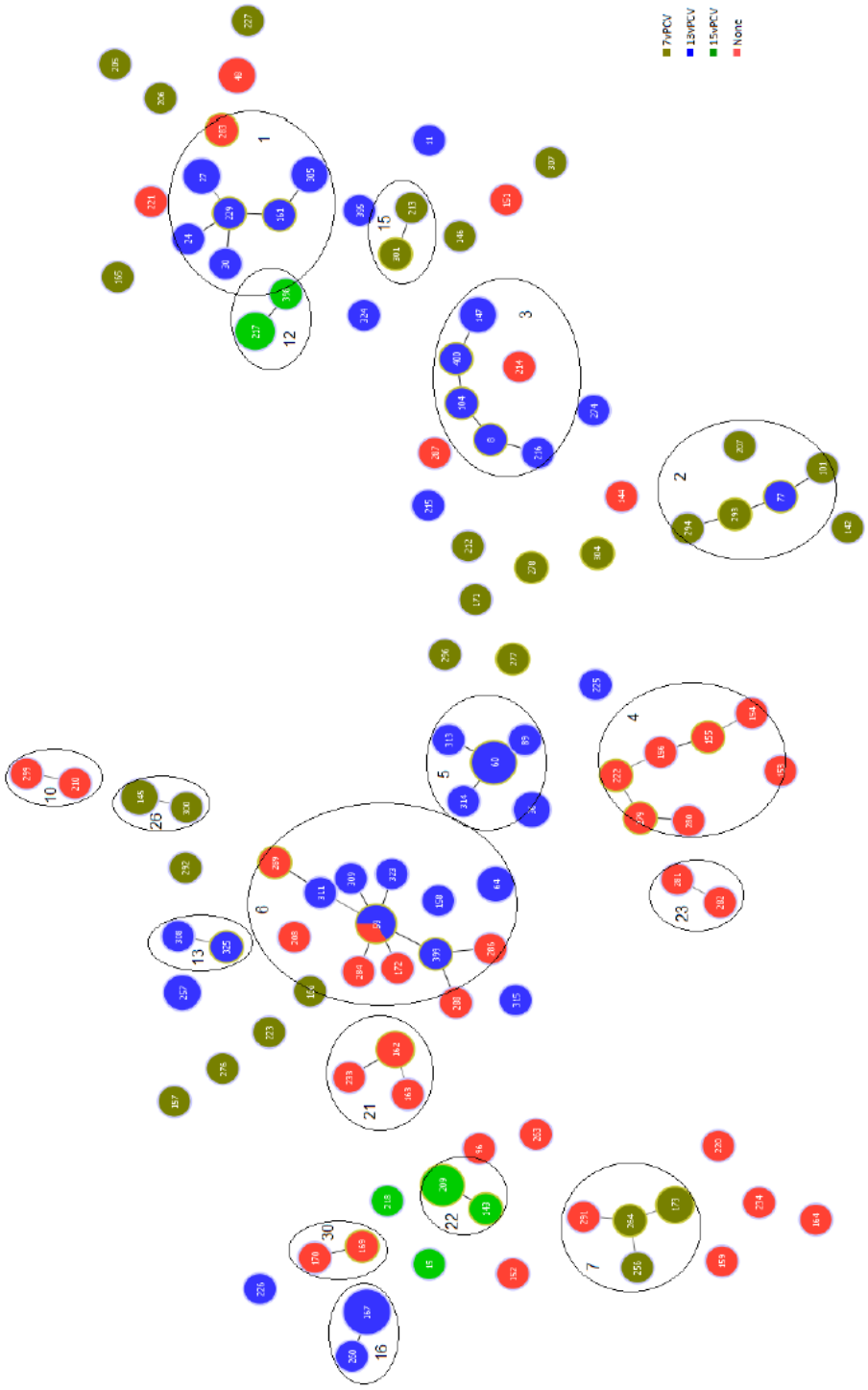
760

761 Figure 3: MLVA4 eBurst of invasive S. pneumoniae isolated from 2007 to
762 2012 in Queensland children with serotypes overlayed (n=317). Clonal
763 clusters (CCs) are circled and contain single locus variants (SLV) and double
764 locus variants (DLV). CCs linked by a thin black line indicate triple locus
765 variants (TLV). MLVA4 genotype is presented inside each circle, and size of
766 circle represents the number of isolates detected in this study. Each colour
767 represents a different serotype, as shown in the legend key. Circles with split
768 colours indicate that two or more serotypes have the same MLVA4 genotype,
769 indicating potential capsule switching.


Key
serotype 1
serotype 3
serotype 4
serotype 6A
serotype 6B
serotype 6C
serotype 7F
serotype 8
serotype 9N
serotype 9V
serotype 10A
serotype 10F
serotype 11A
serotype 12F
serotype 14
serotype 15B
serotype 15C
serotype 16F
serotype 18A
serotype 18B
serotype 18C
serotype 19A
serotype 19F
serotype 22A
serotype 22F
serotype 23A
serotype 23B
serotype 23F
serotype 24F
serotype 33B
serotype 33F
serotype 34
serotype 35B
serotype 35F
serotype 38
CC1
CC2
CC3
CC4
CC6
CC7
CC8
CC9
CC10
CC13

770

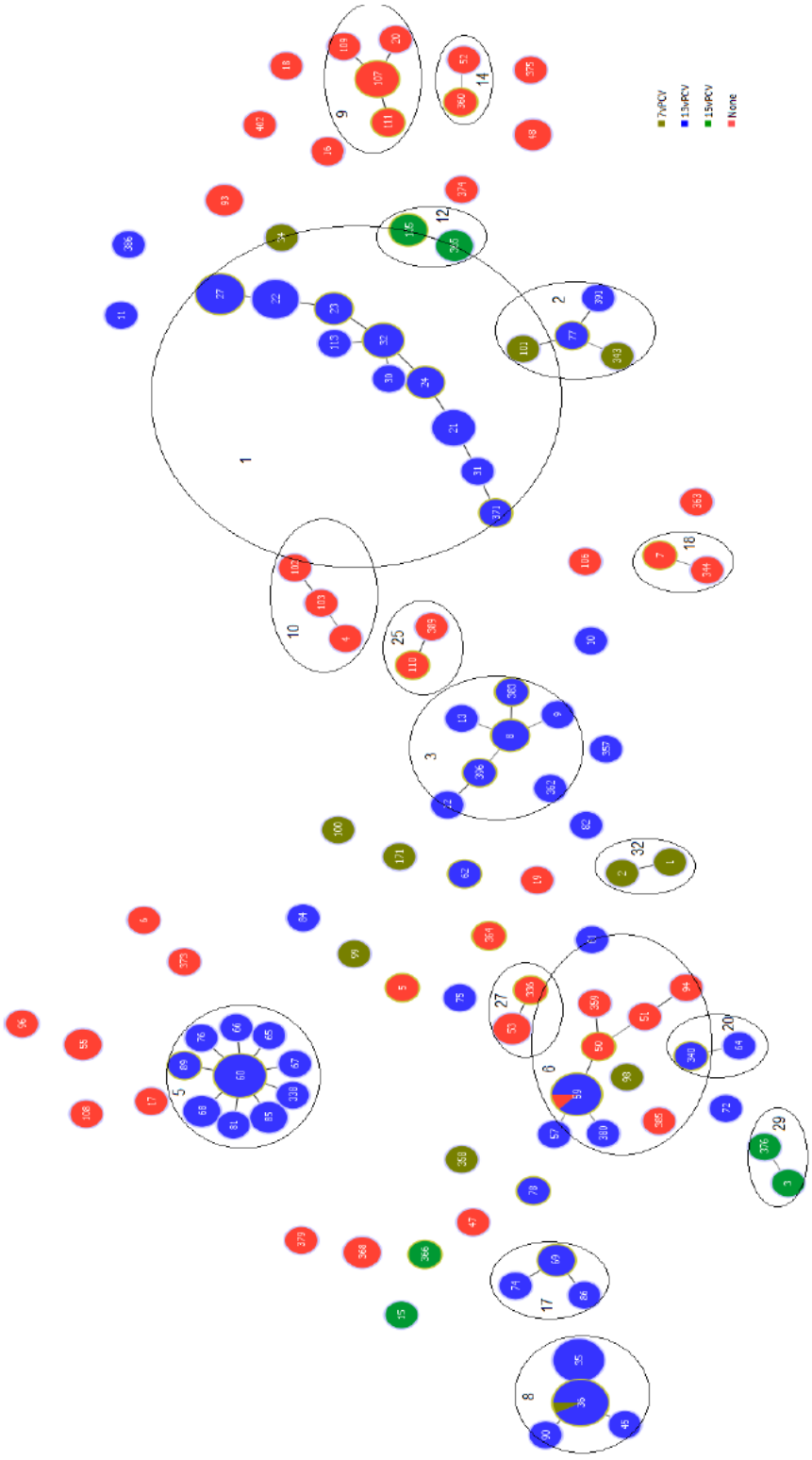
771 Figure 4: MLVA4 eBurst diagram of Queensland invasive S. pneumoniae from
772 2007 to 2009 with designated pneumococcal conjugate vaccine (PCV)
773 serotypes. Clonal clusters (CCs) are circled and contain single locus variants
774 (SLV) and double locus variants (DLV). The 7vPCV serotypes include
775 serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV serotypes include
776 the extra serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV serotypes include
777 serotypes 22F and 33F only. Serotypes not targeted by either of these
778 vaccines are represented as 'none'.



779

38

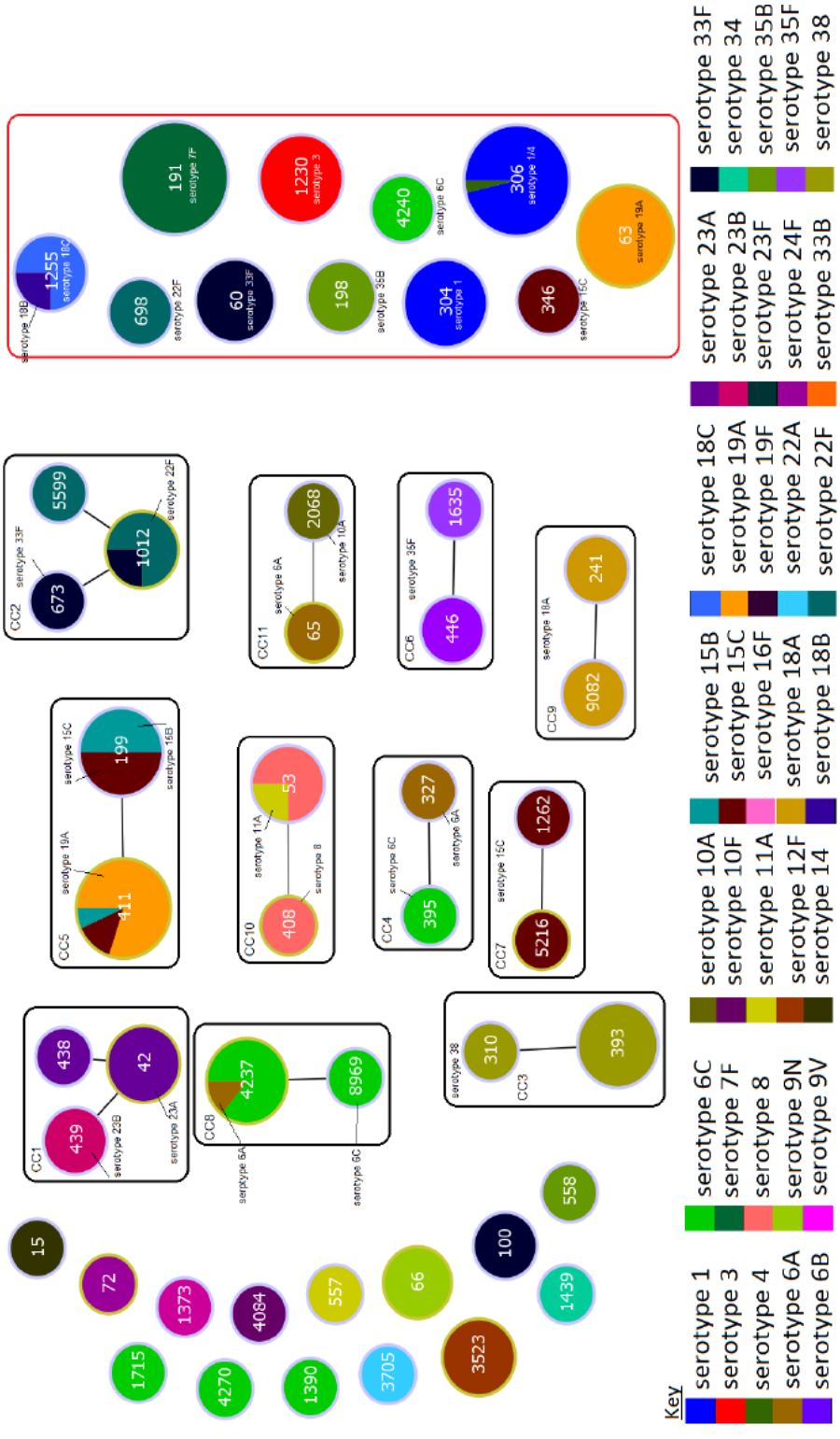
780 Figure 5: MLVA4 eBurst diagram of Queensland invasive S. pneumoniae from
781 2010 to 2012 with designated pneumococcal conjugate vaccine (PCV)
782 serotype. Clonal clusters (CCs) are circled and contain single locus variants
783 (SLV) and double locus variants (DLV). The 7vPCV serotypes include
784 serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. The 13vPCV serotypes include
785 the extra serotypes 1, 3, 5, 6A, 7F and 19A. The 15vPCV serotypes include
786 serotypes 22F and 33F only. Serotypes not targeted by either of these
787 vaccines are represented as 'none'.



788

40

789 Figure 6: MLST eBurst population structure of S. pneumoniae in Queensland
790 from 2007 to 2012 with serotypes overlayed (n=202). Clonal clusters (CCs)
791 are circled and contain single locus variants (SLV) and double locus variants
792 (DLV). MLST sequence type (ST) is presented inside each circle, and size of
793 circle represents the number of isolates detected in this study. Each colour
794 represents a different serotype, as shown in the legend key. Circles with split
795 colours indicate that two or more serotypes have the same MLST genotype,
796 indicating potential capsule switching. Circles within the red box indicate
797 isolates that appear to be singletons (i.e. not genetically related to any other
798 ST within this population study) when using MLST genotyping, but become
799 CCs when using any of the MLVA methods.


Key
serotype 1
serotype 3
serotype 4
serotype 6A
serotype 6B
serotype 6C
serotype 7F
serotype 8
serotype 9N
serotype 9V
serotype 10A
serotype 10F
serotype 11A
serotype 12F
serotype 14
serotype 15B
serotype 15C
serotype 16F
serotype 18A
serotype 18B
serotype 18C
serotype 19A
serotype 19F
serotype 22A
serotype 22F
serotype 23A
serotype 23B
serotype 23F
serotype 24F
serotype 33B
serotype 33F
serotype 34
serotype 35B
serotype 35F
serotype 38
CC1
438
42
439
serotype 23B
serotype 23A
CC2
serotype 33F
5599
673
1012
serotype 22F
CC3
serotype 38
310
393
CC4
serotype 6C
327
395
serotype 6A
CC5
serotype 19A
199
411
serotype 15C
serotype 16B
CC6
serotype 35F
1635
446
CC7
serotype 15C
1262
5216
CC8
4237
serotype 6A
8969
serotype 6C
CC9
serotype 18A
241
9082
CC10
serotype 11A
53
408
serotype 8
CC11
serotype 6A
2068
65
serotype 10A
15
72
1373
4084
557
66
100
558
1715
4270
1390
3705
3523
1439
serotype 18B
1255
serotype 18C
698
serotype 22F
191
serotype 7F
60
serotype 33F
1230
serotype 3
198
serotype 35B
4240
serotype 6C
304
serotype 1
306
serotype 1/4
346
serotype 15C
63
serotype 19A

800

42